献给 A. H. 柯尔莫戈洛夫诞辰一百周年 (1903 — 1987)

目　　录

第一章　初等概率论

我们把只涉及有限个事件概率的, 那一部分概率论称做**初等概率论**.

A. H. 柯尔莫戈洛夫,《概率论的基本概念》[32]

---

§1. 有限种结局试验的概率模型

1. 基本事件空间① 考虑某项试验, 其结果在某一组条件 ("条件复形") 下由有限种不同的结局 (现象) $\omega_1, \cdots, \omega_N$ 描绘. 关于这些结局的实际本性并不重要, 重要的是不同结局的个数 N 是有限的. 我们把这些结局 $\omega_1, \cdots, \omega_N$ 称做*基本事件*, 而把一切结局的全体

$$\Omega = \{\omega_1, \cdots, \omega_N\}$$

称做 (有限) *基本事件空间*, 或*样本空间*.

引进基本事件空间的概念, 是形成不同试验的*概率模型* (概率的理论) 的第一步. 下面是描绘基本事件空间构造的几个例子.

例 1 对于 "掷一枚硬币", 基本事件空间由两个点组成:

$$\Omega = \{\mathrm{Z}, \mathrm{F}\},$$

其中 Z 表示出现 "正面", 而 F 表示出现 "反面". (这时假设, 诸如 "硬币在棱上立着", "硬币丢失" …… 的情况不会出现. 假设不出现 "正面" 就出现 "反面".)

例 2 将一枚硬币重复掷 n 次, 基本事件空间为

$$\Omega = \{\omega : \omega = (a_1, \cdots, a_n)\}, \quad a_i = \mathrm{Z}\text{或 F},$$

且基本事件的总数 $N(\Omega) = 2^n$.

例 3 首先掷一枚硬币. 如果硬币掷出正面 Z, 则掷六面分别刻有 1,2,3,4,5,6 个点的色子; 假如硬币掷出反面 F, 则再掷一次硬币. 那么, 该试验的基本事件空间为

$$\Omega = \{\mathrm{Z1}, \mathrm{Z2}, \mathrm{Z3}, \mathrm{Z4}, \mathrm{Z5}, \mathrm{Z6}, \mathrm{FZ}, \mathrm{FF}\},$$

其中 Z 表示硬币掷出 "正面", F 表示硬币掷出 "反面", 而 1,2,3,4,5,6 表示色子掷出的点数.

---

①原书个别 "小节" 有 "标题". 为便于读者使用, 在没有标题的 "小节", 我们特别加上了标题. —— 译者

注 在讲概率论时, 一般不提 “试验在一组条件下进行”, 而默认这样的 “条件” 存在. 然而, 这样的 “条件” 即便对于初等概率论也很重要, 不同的 “条件” 对同一试验, 也可导致完全不同的概率模型和概率的理论. (参见下文的第 3 小节 “事件的关系和运算” 开头关于这个问题的说明.)

2. 随机抽样与随机分配 现在研究更复杂一些的例子, 这些例子讨论自含 M 个不同球的箱子中, 抽取 n 个球的不同方法.

例 4 *放回抽样*. 如果每次将抽到的球在下一次抽球前放回箱中, 则试验称做放回抽样. 这时, 由 n 个球形成的每一个*样本*可以表示为向量 $(a_1, \cdots, a_n)$, 其中 $a_i (i = 1, 2, \cdots, n)$ 是第 i 次抽到的球的编号. 易见, 对于放回抽样, 每个 a_i 可以是 $1, 2, \cdots, M$ 等 M 个数中任何一个数. 基本事件空间的描述, 本质上与如下情形有关: 诸如 (4,1,2,1) 和 (1,4,2,1) 是认为是不同的基本事件, 还是同一基本事件. 因此, 习惯上区分两种情形: *有序抽样*和*无序抽样*. 在前一种情形下, 由同一些元素组成的两个样本, 只要其中元素的前后顺序有所不同, 就视为*不同的*样本. 在后一种情形下, 不管元素的顺序, 只要由同一些元素组成的样本, 都视为*同一个*样本. 为强调具体的样本究竟是哪一种, 对于*有序样本*, 我们将使用记号 $(a_1, \cdots, a_n)$, 而*无序样本*则记作 $[a_1, \cdots, a_n]$.

于是, *放回有序样本*的基本事件空间 Ω 具有如下构造:

$$\Omega = \{\omega : \omega = (a_1, \cdots, a_n), a_i = 1, 2, \cdots, M\},$$

且不同结局 (样本) 的个数, 等于

$$N(\Omega) = M^n, \tag{1}$$

即组合分析中的自 M 个元素中选 n 个且*允许重复的置换*的总数.

放回无序样本. 即排列组合中的 “*自 M 个元素中选 n 个且允许重复的*组合的总数”, 其基本事件空间为:

$$\Omega = \{\omega : \omega = [a_1, \cdots, a_n], a_i = 1, 2, \cdots, M\}.$$

易见, 不同的无序样本总数 $N(\Omega)$, 小于有序样本总数. 现在证明, 不同无序样本总数:

$$N(\Omega) = \mathrm{C}_{M+n-1}^n, \tag{2}$$

其中 C_k^l *是自 k 个元素中取 l 个的*组合数:

$$\mathrm{C}_k^l = \frac{k!}{l!(k-l)!}.$$

我们用数学归纳法证明. 以 $N(M, n)$ 表示我们感兴趣的结局的个数. 易见, 对于一切 $k \leqslant M$,

$$N(k, 1) = k = \mathrm{C}_k^1.$$

现在假设 $N(k,n)=\mathrm{C}_{k+n-1}^{n}(k\leqslant M)$, 并且证明, 如果将 n 换成 $n+1$ 此式仍然成立. 对于无序样本 $[a_1,\cdots,a_n]$, 可以认为其顺序按不减顺序排列: $a_1\leqslant a_2\leqslant\cdots\leqslant a_{n+1}$. 显然, 无序样本的个数, 当 $a_1=1$ 时为 $N(M,n)$; 当 $a_1=2$ 时为 $N(M-1,n)$; 依次类推, 得

$$\begin{aligned}N(M,n+1)&=N(M,n)+N(M-1,n)+\cdots+N(1,n)\\&=\mathrm{C}_{M+n-1}^{n}+\mathrm{C}_{M-1+n-1}^{n}+\cdots+\mathrm{C}_{n}^{n}\\&=(\mathrm{C}_{M+n}^{n+1}-\mathrm{C}_{M+n-1}^{n+1})+(\mathrm{C}_{M-1+n}^{n+1}-\mathrm{C}_{M-1+n-1}^{n+1})\\&\quad+\cdots+(\mathrm{C}_{n+1}^{n+1}-\mathrm{C}_{n}^{n})+\mathrm{C}_{n}^{n}=\mathrm{C}_{M+n}^{n+1}.\end{aligned}$$

其中用到二项式系数的如下性质:

$$\mathrm{C}_{k}^{l-1}+\mathrm{C}_{k}^{l}=\mathrm{C}_{k+1}^{l}.$$

这一等式, 是由帕斯卡 (B. Pascal) 三角形计算二项式系数的基础, 很容易证明.

例 5 不放回抽样. 假设 $n\leqslant M$, 且凡是抽到的球都不再放回. 这里, 也考虑两种可能的抽法: 有序抽样和无序抽样.

对于不放回有序抽样 (组合分析中称做自 M 个元素中选 n 个的有序无重复排列), 其基本事件空间为

$$\Omega=\{\omega:\omega=(a_1,\cdots,a_n),a_k\neq a_l,k\neq l,a_i=1,2,\cdots,M\},$$

而 Ω 中的元素的总数等于 $M(M-1)\cdots(M-n+1)$. 将此数记作 $(M)_n$ 或 A_M^n, 称做 "自 M 个元素中选 n 个的排列数".

对于不放回无序抽样 (组合分析中称做自 M 个元素中选 n 个无重复的置换), 其基本事件空间为

$$\Omega=\{\omega:\omega=[a_1,\cdots,a_n],a_k\neq a_l,k\neq l,a_i=1,2,\cdots,M\},$$

共含

$$N(\Omega)=\mathrm{C}_M^n \tag{3}$$

个元素. 事实上, 由 n 个不同元素形成的无序数组 $[a_1,\cdots,a_n]$, 可以得到恰好 $n!$ 个有序数组, 从而

$$N(\Omega)\times n!=(M)_n,\quad N(\Omega)=\frac{(M)_n}{n!}=\mathrm{C}_M^n.$$

如果自含 M 个球的箱中选 n 个球, 则可以将不同结局的个数归纳为表 1–1:

表 1－1

	抽样方式		不同抽法的总数
自含 M 个球箱子 Ω 的 n 次抽样	放回	有序	M^n
		无序	C_{M+n-1}^n
	不放回	有序	$(M)_n$
		无序	C_M^n

对于 $M=3, n=2$, 相应基本事件空间的结构列为表 1－2:

表 1－2

	抽样方式		基本事件
自含 M 个球箱子 Ω 的 n 次抽样	放回	有序	$(1,1)(1,2)(1,3)(2,1)(2,2)(2,3)(3,1)(3,2)(3,3)$
		无序	$[1,1][2,2][3,3][1,2][1,3][2,3]$
	不放回	有序	$(1,2)(1,3)(2,1)(2,3)(3,1)(3,2)$
		无序	$[1,2][1,3][2,3]$

例 6 *按箱分配质点.* 考虑按箱分配质点问题: 将 n 个质点 (小球) 分配到 M 个箱中去, 并研究质点分配问题中基本事件空间的构造. 问题产生于统计物理, 例如, 在研究 n 个粒子 (如质子、电子、$\cdots\cdots$) 按 M 种状态的分布.

假设 M 个箱子一一编号为 $1,2,\cdots,M$, 而 n 个质点可以辨别 (两两不同), 且分别编号 $1,2,\cdots,n$. 那么,n 个质点分配到 M 个箱子中, 完全可以由数组 $(a_1,\cdots,a_n)$ 描绘, 其中 a_i 表示第 i 个质点编号为 a_i 的箱中; 假如 n 个质点不可辨别 (完全相同), 则它们分配到 M 个箱中完全可以由数组 $[a_1,\cdots,a_n]$ 描绘, 其中 a_i 表示第 i 个质点分到编号为 a_i 的箱中.

比较例 4 和例 5 所讨论的情形, 可以得到如下对应关系:

(有序抽样) $\Leftrightarrow$ *(质点可以辨)*,

(无序抽样) $\Leftrightarrow$ *(质点不可辨)*.

这种对应关系表示, 在 “自盛有 M 个球的箱子中抽选 n 个球” 的问题中, 有序 (无序) 抽样相当于, 可辨质点 (不可辨质点) 分配到 M 个箱中的分配问题.

下面的对应关系具有类似的含义:

(放回抽样) $\Leftrightarrow$ *(箱中可容纳任意多质点)*,

(不放回抽样) $\Leftrightarrow$ *(箱中最多可容纳一个质点)*.

由这些对应关系, 可以建立如下类型的对应关系:

(无序不放回抽样) $\Leftrightarrow$ *(不可辨质点按箱分配, 且每箱最多容纳一个质点)*

等等. 由于这些对应关系, 就可以在可辨和不可辨质点按箱分配问题中, 利用例 4 和例 5 描绘基本事件空间, 既包括 “有抑制” 的箱子 (最多容纳一个质点的箱子) 和 “无抑制” 的箱子 (可容纳任意个质点的箱子).

表 1–3 演示按两个箱子分配 3 个小球的情形, 其中 “○” 和 “●” 分别表示白球和黑球. 如果分到箱中的球不可辨, 则用 “+” 表示.

表 1 – 3

	分配方式		基本事件
自含 M 个球箱子 Ω 中的 n 次抽样	有抑制	可辨	[●| |○] [○| |●] [○|●|] [●|○|] [|○|●] [|●|○]
		不可辨	[+| |+] [|+|+] [+|+|]
	无抑制	可辨	[○| |●] [●|○|] [|○|●] [| |○●] [|○●|] [●| |○] [○|●|] [|●|○] [○●| |]
		不可辨	[++| |] [|++|] [| |++] [+|+|] [+| |+] [|+|+]

对于质点按箱分配问题中, 利用上面指出的 “抽样” 与 “分配” 之间的对偶性, 就可以明显地求出分配结局的个数. 相应的结果归纳为表 1–4, 其中也包含表 1 – 1 的结果.

表 1 – 4

	抽样方式		不同抽法的总数	分配方式		
自含 M 个球箱子 Ω 中的 n 次抽样	放回	有序	M^n ①	可辨	有抑制	自含 M 个球箱子 Ω 中的 n 次抽样
		无序	C_{M+n-1}^{n} ②	不可辨		
	不放回	有序	$(M)_n$ ③	可辨	无抑制	
		无序	C_{M}^{n} ④	不可辨		
	抽样方式		不同抽法的总数	分配方式		

统计物理学中, 把不满足帕利 (W. Pauli) 抑制原理 (“每箱中最多容纳一个质点”) 的不同 (不可辨) 质点, 称做麦克斯韦 – 波尔茨曼 (J. C. Maxwell-L. Boltzmann) (物理) 统计 (见表 1–4 ①); 有抑制不可辨质点的情形称做博泽 – 爱因斯坦 [S. N. Bose-A. Einstein] 统计 (见表 1–4 ②). 如果质点不可辨但满足抑制原理, 则称做弗米 – 迪拉克 (E. Fermi-P. A. M. Dirac) 统计 (见表 1–4 ④). 熟知, 例如电子、质子和中子都服从弗米 – 迪拉克统计, 光子和 π 介子服从博泽 – 爱因斯坦统计 (见表 1–4 ②). 此外, 在物理学中再没有发现可辨粒子不服从抑制原理的情形.

3. 事件及其关系和运算　除基本事件空间的概念外, 现在引进重要概念 —— 事件. 事件的概念, 是建立所考察试验的各种概率模型 (“理论”) 的基础. 在试验的结果中, 试验者一般并不关心究竟出现了哪种具体的结局, 而关心出现的结局属于

一切结局集合的哪个子集. 满足试验条件的一切子集 $A \subseteq \Omega$, 分为两种类型: "结局 $\omega \in A$" 或 " 结局 $\omega \notin A$". 我们称这样的子集 A 为*事件*.

例如, 将一枚硬币重复掷三次, 一切可能结局的空间 Ω, 由 8 个点构成:

$$\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\},$$

其中 0 和 1 分别表示掷出 "正面" 和 "反面". 如果由 "一组条件" 可以记录 (确定、测量等) 所有 3 次掷硬币的结果, 则例如

$$A = \{000, 001, 010, 100\}$$

就是事件: "将一枚硬币重复掷三次" 正面至少出现两次. 假如由 "一组条件" 只能确定第一次掷出的结果, 则 A 已经不能称为*事件*, 因为关于 "试验的具体结局 ω 是否属于 A", 既不能肯定也不能否定.

从给定的作为事件的某一组集合出发, 经过某些变换可以形成一些新事件. 这些变换, 可以用逻辑的语言表述为 "或"、"与"、"非", 而用集合论的语言表述为 "并"、"交" 和 "补".

并 对于两个集合 A 与 B, 称

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ 或 } \omega \in B\}$$

为集合 A 与 B 的*并*, 表示由属于集合 A 或 B 的点形成的集合. 用概率论的语言, $A \bigcup B$ 表示事件 {事件 A 与 B 至少出现一个}.

交 称 AB 或

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ 且 } \omega \in B\}$$

为两个集合 A 与 B 的*交*, 表示由既属于集合 A 同时又属于集合 B 的点形成的集合. 用概率论的语言, $A \bigcap B$ 表示事件 {事件 A 与 B 同时出现}.

例如, 如果 $A = \{00, 01, 10\}, B = \{11, 10, 01\}$, 则

$$A \cup B = \{00, 01, 10, 11\} = \Omega, \quad A \cap B = \{10, 01\}.$$

补 如果 A 是 Ω 的子集, 则称 $\overline{A} = \{\omega \in \Omega : \omega \notin A\}$ 为集合 A *补*, 表示 Ω 中不属于 A 的点的集合.

差 属于 B 但不属于 A 的点的集合称做 B 与 A 的*差*, 记作 $B \backslash A$. 那么,$\overline{A} = \Omega \backslash A$. 用概率的语言,$\overline{A}$ 表示事件 "A 不出现". 例如, 事件 $A = \{00, 01, 10\}$, 则事件 $\overline{A} = \{11\}$ 表示接连两次掷出反面.

不可能事件和必然事件 一般用 $\varnothing$ 表示空集. 概率论中空集 $\varnothing$ 称做*不可能事件*, 集合 Ω 自然称做*必然事件*. 集合 A 和 $\overline{A}$ 没有共同点, 因而 $A \bigcap \overline{A}$ 是空集.[①]

---

①对于事件 A 和 B, 若 $A \bigcap B = \varnothing$, 则称 A 和 B 不相容, 否则称 A 和 B 相容. —— 译者

和　当 A 与 B 不相交时 $(AB=\varnothing)$, 集合 A 与 B 的并称做集合 A 与 B 的**和**, 记作 $A+B$.

事件代数　考虑集合 $A\subseteq\Omega$ 的某个集系 $\mathscr{A}_0$, 则利用集合运算 $\cup,\cap$ 与 $\setminus$ 可以由 $\mathscr{A}_0$ 构造新集系, 其中元素也是事件. 给这些事件补充上必然事件 Ω 和不可能事件 $\varnothing$, 得集系 $\mathscr{A}$, 则 $\mathscr{A}$ 是代数. 所谓 "代数" 即 Ω 的这样的集系, 满足

1) $\Omega\in\mathscr{A}$,

2) 若 $A\in\mathscr{A},B\in\mathscr{A}$, 则集合 $A\cup B,A\cap B,A\backslash B$ 也都属于 $\mathscr{A}$.

由以上的叙述, 可见作为事件系最好考虑本身是代数的集系. 以后, 我们正是考虑这样的集系.

我们列举几个事件代数的例子.

a) $\mathscr{A}=\{\Omega,\varnothing\}$ —— 集系由 Ω 和空集 $\varnothing$ 构成, 称做*平凡代数*;

b) $\mathscr{A}=\{A,\overline{A},\Omega,\varnothing\}$ —— 事件 A 产生的集系;

c) $\mathscr{A}=\{A:A\subseteq\Omega\}$ —— Ω 全部子集的集系 (包括空集 $\varnothing$).

易见, 所有这些事件代数是按下面的原则得到的.

分割　我们称集系

$$\mathscr{D}=\{D_1,\cdots,D_n\}$$

构成集合 Ω 的一个*分割*, 而 $D_1,\cdots,D_n$ 是该分割的*原子*, 如果 $D_1,\cdots,D_n$ 非空且两两不相容, 而它们的和等于 Ω:

$$D_1+\cdots+D_n=\Omega.$$

例如, 假定集合 Ω 由 3 个点构成: $\Omega=\{1,2,3\}$, 则存在 5 个不同的分割:

$$\begin{array}{ll}
\mathscr{D}_1=\{D_1\} & D_1=\{1,2,3\};\\
\mathscr{D}_2=\{D_1,D_2\} & D_1=\{1,2\},D_2=\{3\};\\
\mathscr{D}_3=\{D_1,D_2\} & D_1=\{1,3\},D_2=\{2\};\\
\mathscr{D}_4=\{D_1,D_2\} & D_1=\{2,3\},D_2=\{1\};\\
\mathscr{D}_5=\{D_1,D_2,D_3\} & D_1=\{1\},D_2=\{2\};D_3=\{3\}.
\end{array}$$

(关于有限集合的分割的总数参见本节的习题 2.)

如果考虑 $\mathscr{D}$ 中一切集合的并连同空集 $\varnothing$, 则得到的集系是代数, 称做 *$\mathscr{D}$ 产生的代数*, 记作 $\sigma(\mathscr{D})$. 于是, 代数 $\sigma(\mathscr{D})$ 的元素由空集 $\varnothing$ 与分割 $\mathscr{D}$ 之原子中集合的和组成.

这样, 如果 $\mathscr{D}$ 是 Ω 的某一分割, 则它与代数 $\mathscr{B}=\sigma(\mathscr{D})$ 一一对应.

逆命题也正确. 设 $\mathscr{B}$ 是有限空间 Ω 的子集的代数, 则存在唯一分割 $\mathscr{D}$, 其原子是代数 $\mathscr{B}$ 的元素, 并且 $\mathscr{B}=\sigma(\mathscr{D})$. 事实上, 假设集合 $\mathscr{D}\in\mathscr{B}$ 并且具有性质: 对于任意 $B\in\mathscr{B}$, 集合 $D\bigcap B$ 要么与 D 重合, 要么是空集. 那么, 这样集合 D 的全体组成分割 $\mathscr{D}$ 并且具有所要求的性质 $B=\sigma(\mathscr{D})$. 对于例 a), 取只含一个集合

的 $D_1 = \Omega$ 平凡分割; 对于例 b), $\mathscr{D} = \{A, \overline{A}\}$. 对于例 c), $\mathscr{D}$ 是只含一个点的集合 $\{\omega_i\}, \omega_i \in \Omega$ 的最细小分割, 即 $\mathscr{D}$ 产生的代数是 Ω 的一切子集的代数.

对两个分割 $\mathscr{D}_1$ 和 $\mathscr{D}_2$, 如果 $\sigma(\mathscr{D}_1) \subseteq \sigma(\mathscr{D}_2)$, 则称分割 $\mathscr{D}_2$ 比分割 $\mathscr{D}_1$ "细小", 记作 $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2$.

像前面一样, 假设空间 Ω 有有限个点 $\omega_1, \cdots, \omega_N$ 构成, 记 $N(\mathscr{A})$ 为例 c) 中组成体系 $\mathscr{A}$ 的集合的总数. 我们证明 $N(\mathscr{A}) = 2^N$. 事实上, 每一个非空集合 $A \in \mathscr{A}$ 可以表示为 $A = \{\omega_{i_1}, \cdots, \omega_{i_k}\}(1 \leqslant k \leqslant N)$, 其中 $\omega_{i_j} \in \Omega$. 将该集合与由 0 或 1 形成的序列

$$(0, \cdots, 0, 1, 0, \cdots, 0, 1, \cdots),$$

其中在编号为 $i_1, \cdots, i_k$ 的位置上为 1, 而在其余位置上为 0. 那么, 对于固定的 k, 形如 $A = \{\omega_{i_1}, \cdots, \omega_{i_k}\}$ 的不同集合 A 的总数等于 k 个 1 (k 个不可辨质点) 分配 N 个位置 (N 个箱子) 不同分法的总数. 根据表 1 – 4 的情形④, 这样分法的总数等于 C_N^k. 由此可见

$$N(\mathscr{A}) = 1 + \mathrm{C}_N^1 + \cdots + \mathrm{C}_N^N = 2^N,$$

其中包括空集 $\varnothing$.

4. 概率空间 为建立只有有限种可能结局的随机试验的概率模型 (理论), 我们暂时完成了最初的两步: 引进了基本事件空间 Ω, 并建立了 Ω 子集的某种体系 $\mathscr{A}$ —— 代数, 其中的子集称做事件. 有时把 $\mathscr{E} = (\Omega, \mathscr{A})$ 等同于试验. 现在进行下一步: 赋予每一个基本事件 (每一种结局或现象) $\omega_i \in \Omega (i = 1, \cdots, N)$ 某种 "权", 记作 $p(\omega_i)$ 或 p_i, 称做基本事件 (结局) ω_i 的概率. 假设 $p(\omega_i)$ 满足条件:

a) $0 \leqslant p(\omega_i) \leqslant 1$ (非负性),

b) $p(\omega_1) + \cdots + p(\omega_N) = 1$ (规范性).

从给定的基本事件 ω_i 的概率 $p(\omega_i)$ 出发, 按公式

$$\mathbf{P}(A) = \sum_{\{i: \omega_i \in A\}} p(\omega_i) \tag{4}$$

定义任意事件 $A \in \mathscr{A}$ 的概率.

定义 1 通常称

$$(\Omega, \mathscr{A}, \mathbf{P})$$

为 "概率空间", 其中 $\Omega = \{\omega_1, \cdots, \omega_N\}$, $\mathscr{A}$ 是 Ω 的子集的代数, 而 $\mathbf{P} = \{\mathbf{P}(A) : A \in \mathscr{A}\}$. 概率空间决定 (定义) 只有有限种可能结局 (基本事件) 的, 随机试验的概率模型 (理论). 显然, $\mathbf{P}(\{\omega_i\}) = p(\omega_i) (i = 1, \cdots, N)$

$$\mathbf{P}(\varnothing) = 0, \tag{5}$$

$$\mathbf{P}(\Omega) = 1, \tag{6}$$

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B). \tag{7}$$

特别, 若 $A \cap B = \varnothing$, 则

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B), \tag{8}$$

而

$$\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A), \tag{9}$$

5. 古典型概率 在一些具体的情形下建立概率模型时, 给出基本事件空间 Ω 和代数 $\mathscr{A}$, 一般并不复杂. 这时, 在初等概率论里, 一般用 Ω 的全部子集的代数当做代数 $\mathscr{A}$. 较困难的问题, 是定义基本事件的概率. 实际上, 对这个问题的回答已经超出了概率论的范围, 我们不在此过多地讨论这个问题. 我们的基本任务, 并不是如何赋予某一个试验基本事件的概率, 而是根据基本事件的概率, 计算复合事件 ($\mathscr{A}$ 中的事件) 的概率.

从数学观点来看十分清楚, 对于有限基本事件空间, 只要赋予基本事件 $\omega_1, \cdots, \omega_N$ 以满足 $p_1 + \cdots + p_N = 1$ 的非负实数 $p_1, \cdots, p_N$, 就可以得到一切可以想象的 (有限的) 概率空间.

对于具体的情形, 所确定的数值的*正确性*, 可以一定程度地利用以后将要介绍的*大数定律*来验证. 在上述情形下, 根据大数定律, 对于给定的在相同条件下进行的较长 "独立" 试验系列, 基本事件出现的频率 "十分接近" 它们相应的概率.

鉴于赋予试验基本事件以概率值的困难, 我们指出, 存在许多实际情形, 在这些情形下由于*对称性*或*均衡性*的直观, 把一切可能出现的基本事件视为*等可能的*是合理的. 因此, 假如基本事件空间 Ω 由点 $\omega_1, \cdots, \omega_N$ 构成, 其中 $N < \infty$, 则

$$p(\omega_1) = \cdots = p(\omega_N) = \frac{1}{N},$$

从而对于任何事件 $A \in \mathscr{A}$,

$$\mathbf{P}(A) = \frac{N(A)}{N}, \tag{10}$$

其中 $N(A)$ 是事件 A 所含基本事件的个数.

这样求概率的方法称做*古典型方法*. 显然, 这时求概率 $\mathbf{P}(A)$ 归结为计算导致事件 A 的基本事件的*个数*. 这一般用*排列组合*的方法来实现. 因此, 在涉及有限集合的概率的计算中, 排列组合的方法占有重要的地位.

例 7 *重合问题*. 假设箱中有 M 个球, 分别编号为 $1, 2, \cdots, M$. 进行 n 次放回抽样, 并且认为样本是有序的. 显然, 这时

$$\Omega = \{\omega : \omega = (a_1, \cdots, a_n), a_i = 1, \cdots, M\},$$

而 $N(\Omega) = M^n$. 根据古典型概率的求法, 认为所有 M^n 个结局是等可能的. 提出下面的问题: 事件

$$A = \{\omega : a_i \neq a_j, i \neq j\}$$

的概率如何? 即事件 "抽到的元素无重复" 的概率如何? 显然

$$N(A) = M(M-1)\cdots(M-n+1) = (M)_n,$$

因此

$$\mathbf{P}(A) = \frac{(M)_n}{M^n} = \left(1-\frac{1}{M}\right)\left(1-\frac{2}{M}\right)\cdots\left(1-\frac{n-1}{M}\right). \tag{11}$$

该例题有如下有趣的解释. 假设某一班中有 n 名学生, 每个学生的生日在一年 365 天中每一天是等可能的. 问 "n 个学生中至少有两人的生日在同一天" 的概率 $P_{365}(n)$ 如何? 可以把 "选择" 生日, 视为在 365 个球中任选一个球, 则根据 (11) 式

$$P_{365}(n) = 1 - \frac{(365)_n}{365^n}.$$

对于某些 n_3 下表列出了概率 $P_{365}(n)$ 的值:

n	4	16	22	23	40	64	70	100
$P_{365}(n)$	0.016 36	0.283 60	0.475 69	0.507 30	0.891 23	0.997 11	0.999 16	0.999 999 69

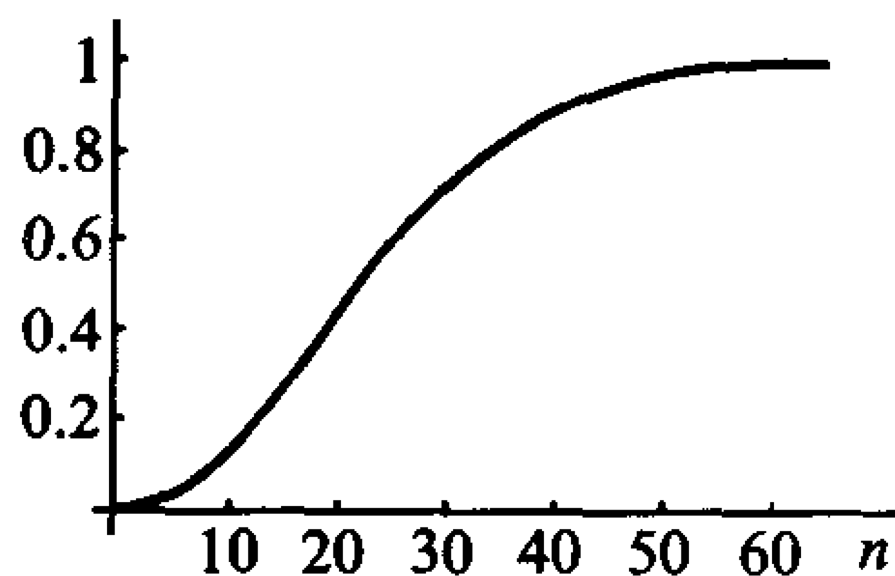
对于充分大的 M,

$$\ln\frac{(M)_n}{M^n} = \sum_{k=1}^{n-1}\ln\left(1-\frac{k}{M}\right) \sim -\frac{1}{M}\sum_{k=1}^{n-1}k = -\frac{n(n-1)}{2M},$$

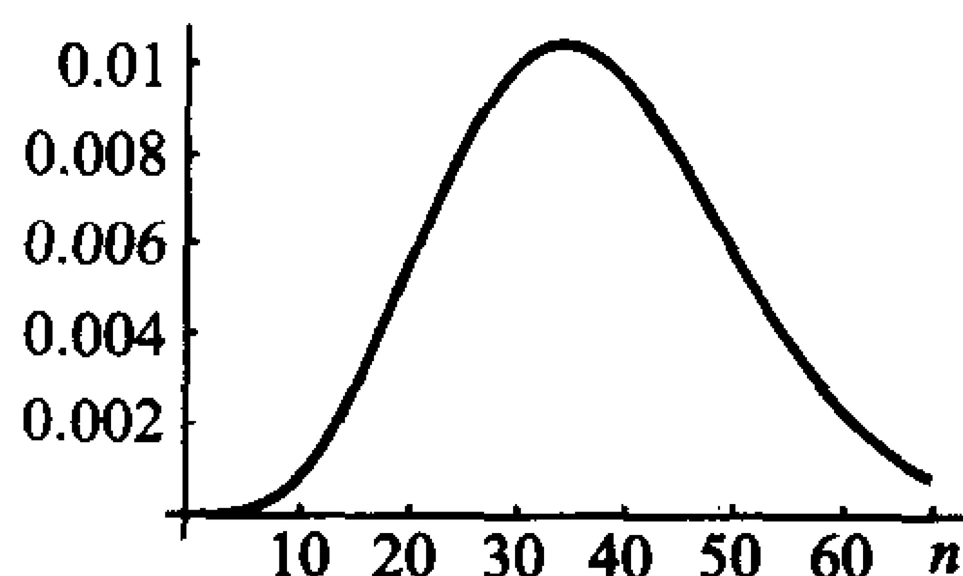
亦即

$$P_M(n) \equiv 1 - \frac{(M)_n}{M^n} \sim 1 - \mathrm{e}^{-\frac{n(n-1)}{2M}} \quad [\equiv \widetilde{P}_M(n)], \quad M \to \infty.$$

下面是函数 $P_{365}(n)$ 的图形. 在相同的尺度下, $\widetilde{P}_{365}(n)$ 逼近的图形实际上与函数 $P_{365}(n)$ 的图形相重合. 在区间 $[0,60]$ 的范围内二者的最大差别近似等于 0.01(在 $n=30$ 的邻域内).



$P_{365}(n)$ 和 $\widetilde{P}_{365}(n)$ 图形



$P_{365}(n) - \widetilde{P}_{365}(n)$ 图形

有趣的是, (和期望的相反!) 当一个班的人数并不是太多, 只有 23 名学生时, "n 个学生中至少有两人的生日在同一天" 的概率就已达 1/2.

例 8 **抽彩**. 考虑按如下规则的抽彩. 假设总共有 M 张编号为 $1,2,\cdots,M$ 彩票, 其中编号为 $1,2,\cdots,n(M \geqslant 2n)$ 的中彩. 一人买了 n 张彩票, 问至少有一张中奖的概率 (记作 P) 如何?

由于抽取彩票的顺序, 对于所购买的彩票是否中奖无意义, 所以可以认为基本事件空间有如下结构:

$$\Omega = \{\omega : \omega = [a_1, \cdots, a_n], a_k \neq a_l, k \neq l, a_i = 1, \cdots, M\}.$$

由表 1 – 1 知 $N(\Omega) = \mathrm{C}_M^n$. 现在假设

$$A_0 = \{\omega : \omega = [a_1, \cdots, a_n], a_k \neq a_l, k \neq l, a_i = n+1, \cdots, M\}$$

是事件 "所买的彩票中没有中奖的". 仍然由表 1 – 1 知 $N(A_0) = \mathrm{C}_{M-n}^n$. 因此,

$$\mathbf{P}(A_0) = \frac{\mathrm{C}_{M-n}^n}{\mathrm{C}_M^n} = \frac{(M-n)_n}{(M)_n} = \left(1 - \frac{n}{M}\right)\left(1 - \frac{n}{M-1}\right)\cdots\left(1 - \frac{n}{M-n+1}\right)$$

即

$$P = 1 - \mathbf{P}(A_0) = 1 - \left(1 - \frac{n}{M}\right)\left(1 - \frac{n}{M-1}\right)\cdots\left(1 - \frac{n}{M-n+1}\right).$$

如果 $M = n^2$ 且 $n \to \infty$, 则 $\mathbf{P}(A_0) \to \mathrm{e}^{-1}$,

$$P \to 1 - \mathrm{e}^{-1} \approx 0.632,$$

其中收敛的速度相当快: 当 $n = 10$ 时, 概率已经为 $P = 0.670$.

6. 练习题

1. 验证运算 $\bigcap$ 和 $\bigcup$ 的下列性质的正确性:

交换律: $A \bigcup B = B \bigcup A, A \bigcap B = B \bigcap A$;

结合律: $A \bigcup (B \bigcup C) = (A \bigcup B) \bigcup C, A \bigcap (B \bigcap C) = (A \bigcap B) \bigcap C$;

分配律: $A \bigcap (B \bigcup C) = (A \bigcap B) \bigcup (A \bigcap C), A \bigcup (B \bigcap C) = (A \bigcup B) \bigcap (A \bigcup C)$;

幂等性: $A \bigcup A = A, A \bigcap A = A$;

对偶律: $\overline{A \bigcup B} = \overline{B} \bigcap \overline{A}, \overline{A \bigcap B} = \overline{B} \bigcup \overline{A}$.

2. 设集合 Ω 含 N 个元素. 证明不同分割的贝尔 (A.G.Bell) 数 B_N 决定于

$$B_N = \mathrm{e}^{-1} \sum_{k=1}^{\infty} \frac{k^N}{k!}. \tag{12}$$

提示 证明

$$B_N = \sum_{k=0}^{N-1} \mathrm{C}_{N-1}^k B_k, \quad B_0 = 1,$$

然后证明 (12) 式满足上面的递推公式.

3. 证明, 对于任何有限集合系 $A_1, \cdots, A_n$, 有 (概率的半可加性)

$$\mathbf{P}(A_1 \cup \cdots \cup A_n) \leqslant \mathbf{P}(A_1) + \cdots + \mathbf{P}(A_n).$$

4. 设 A 和 B 是任意二事件, 证明, $A\overline{B}\bigcup B\overline{A}$ 表示事件 "A 和 B 中必出现一个且只出现一个", 并且

$$\mathbf{P}(A\overline{B}\cup B\overline{A})=\mathbf{P}(A)+\mathbf{P}(B)-2\mathbf{P}(AB).$$

5. 设有 $A_1,\cdots,A_n$, 而量 $S_0,S_1,\cdots,S_n$ 决定于如下关系式: $S_0=1$,

$$S_r=\sum_{J_r}\mathbf{P}(A_{k_1}\cap\cdots\cap A_{k_r}),\quad 1\leqslant r\leqslant n,$$

其中对集合 $\{1,2,\cdots,n\}$ 的无序子集 $J_r=[k_1,\cdots,k_r](k_i\neq k_j,i\neq j)$ 求和.

(1) 设 B_m 表示事件 "在 $A_1,\cdots,A_n$ 中必出现一个且只出现一个". 证明

$$\mathbf{P}(B_m)=\sum_{r=m}^{n}(-1)^{r-m}\mathrm{C}_r^m S_r.$$

特别, 对于 $m=0$,

$$\mathbf{P}(B_0)=1-S_1+S_2-\cdots\pm S_n$$

(2) 证明, "$A_1,\cdots,A_n$ 中有 m 个同时出现" 的概率等于

$$\mathbf{P}(B_m)+\cdots+\mathbf{P}(B_n)=\sum_{r=m}^{n}(-1)^{r-m}\mathrm{C}_{r-1}^{m-1}S_r.$$

特别, "$A_1,\cdots,A_n$ 中至少出现一个" 的概率等于

$$\mathbf{P}(B_1)+\cdots+\mathbf{P}(B_n)=S_1-S_2+\cdots\mp S_n.$$

(3) 证明下列性质:

(a) 邦弗尔罗尼 (Bonferroni) 不等式[①]: 对于任意 $k=1,2,\cdots,(2k\leqslant n)$, 有.

$$S_1-S_2+\cdots-S_{2k}\leqslant\mathbf{P}\left(\bigcup_{i=1}^{n}A_i\right)=S_1-S_2+\cdots+S_{2k-1};$$

(b) 庞加莱 (J. H. Poincare) 恒等式:

$$\mathbf{P}\left(\bigcup_{r=1}^{n}A_r\right)=\sum_{r=m}^{n}(-1)^{r-1}S_r;$$

(c) 弗雷歇 (M. Fréshet) 不等式:

$$\frac{S_{r+1}}{\mathrm{C}_n^{r+1}}\leqslant\frac{S_r}{\mathrm{C}_n^r},\quad(r=0,1\cdots,n-1);$$

① 见 W. Feller, An Introduction to probability theory, V. I, ch. Ⅳ. —— 译者

(d) 冈贝尔 (E. J. Gumbel) 不等式:

$$\frac{C_n^{r+1}-S_{r+1}}{C_{n-1}^r} \leqslant \frac{C_n^r-S_r}{C_{n-1}^{r-1}}, \quad (r=1,2,\cdots,n-1);$$

6. 证明 $\mathbf{P}(A\bigcap B\bigcap C) \geqslant \mathbf{P}(A)+\mathbf{P}(B)+\mathbf{P}(C)-2$, 并用归纳法证明

$$\mathbf{P}\left(\bigcap_{i=1}^n A_i\right) \geqslant \sum_{i=1}^n \mathbf{P}(A_i)-(n-1)$$

7. 在关于 n 和 M 不同类型 (类型: $n=xM, M\to\infty$ 或 $n=x\sqrt{M}, M\to\infty$, 其中 x 固定) 的假设下, 讨论例 7 中概率 $P_M(n)$ 的渐近性. 将结果与 §6 中的局部极限定理进行比较.

§2. 某些经典模型和分布

1. 二项分布 假设将一枚硬币接连掷 n 次, 观测结果用有序数组 $(a_1,\cdots,a_n)$ 表示, 其中, 当第 i 次掷出现正面时 $a_i=1$, 当第 i 次掷出现反面时 $a_i=0$. 基本事件空间具有如下形式:

$$\Omega=\{\omega:\omega=(a_1,\cdots,a_n), a_i=0 \text{ 或 } 1\}.$$

赋予每一个基本事件 $\omega=(a_1,\cdots,a_n)$ 概率 (权重)

$$p(\omega)=p^{\sum a_i}q^{n-\sum a_i},$$

其中 p 和 q 非负且 $p+q=1$. 首先证明, 这样定义概率 (权重) $p(\omega)$ 是合理的. 为此, 只需验证

$$\sum_{\omega\in\Omega}p(\omega)=1.$$

考虑所有满足

$$\sum_i a_i=k, \quad (k=0,1,\cdots,n)$$

的基本事件 $\omega=(a_1,\cdots,a_n)$. 根据表 1 – 4 (k 个不可辨的 “1” 分配到 n 个位置上), 这样的基本事件个数等于 C_n^k. 因此

$$\sum_{\omega\in\Omega}p(\omega)=\sum_{k=0}^n C_n^k p^k q^{n-k}=(p+q)^n=1.$$

设 $\mathscr{A}$ 是空间 Ω 的一切子集的代数, 在 $\mathscr{A}$ 上定义了概率:

$$\mathbf{P}(A)=\sum_{\omega\in A}p(\omega), \quad A\in\mathscr{A},$$

其中 $\mathbf{P}(\{\omega\}) = p(\omega), \omega \in \Omega$, 从而定义了一个概率模型. 自然称做描绘 n 次掷硬币的概率模型.

当 $n = 1$ 时, 基本事件空间仅含两个点 $\omega = 1$ ("成功") 和 $\omega = 0$ ("失败"), 而概率 $p(1) = p$, 自然称做 "成功的概率". 以后, 我们将看到, 作为 n 次 "独立" 试验的结果, 可以得到由上面讨论的描绘 n 次掷硬币的概率模型, 其中 p 是每次试验成功的概率.

事件

$$A_k = \{\omega : \omega = (a_1, \cdots, a_n), a_1 + \cdots + a_n = k\}, \quad k = 0, 1, \cdots, n,$$

表示恰好 k 次 "成功". 由以上的讨论, 可见

$$\mathbf{P}(A_k) = \mathrm{C}_n^k p^k q^{n-k}, \text{ 且 } \sum_{k=0}^{n} \mathbf{P}(A_k) = 1. \tag{1}$$

概率 $(\mathbf{P}(A_0), \mathbf{P}(A_1), \cdots, \mathbf{P}(A_n))$ 组称做二项分布 (在容量为 n 的样本中 "成功" 的次数服从二项分布).

二项分布在概率论中起着十分重要的作用, 出现在多种不同的概率模型中. 记

$$P_n(k) = \mathbf{P}(A_k), \quad k = 0, 1, \cdots, n.$$

对于 $p = 1/2$ (掷 "对称" 硬币) 的情形, 当 $n = 5, 10, 20$ 时, 图 1 是二项分布的示意图.

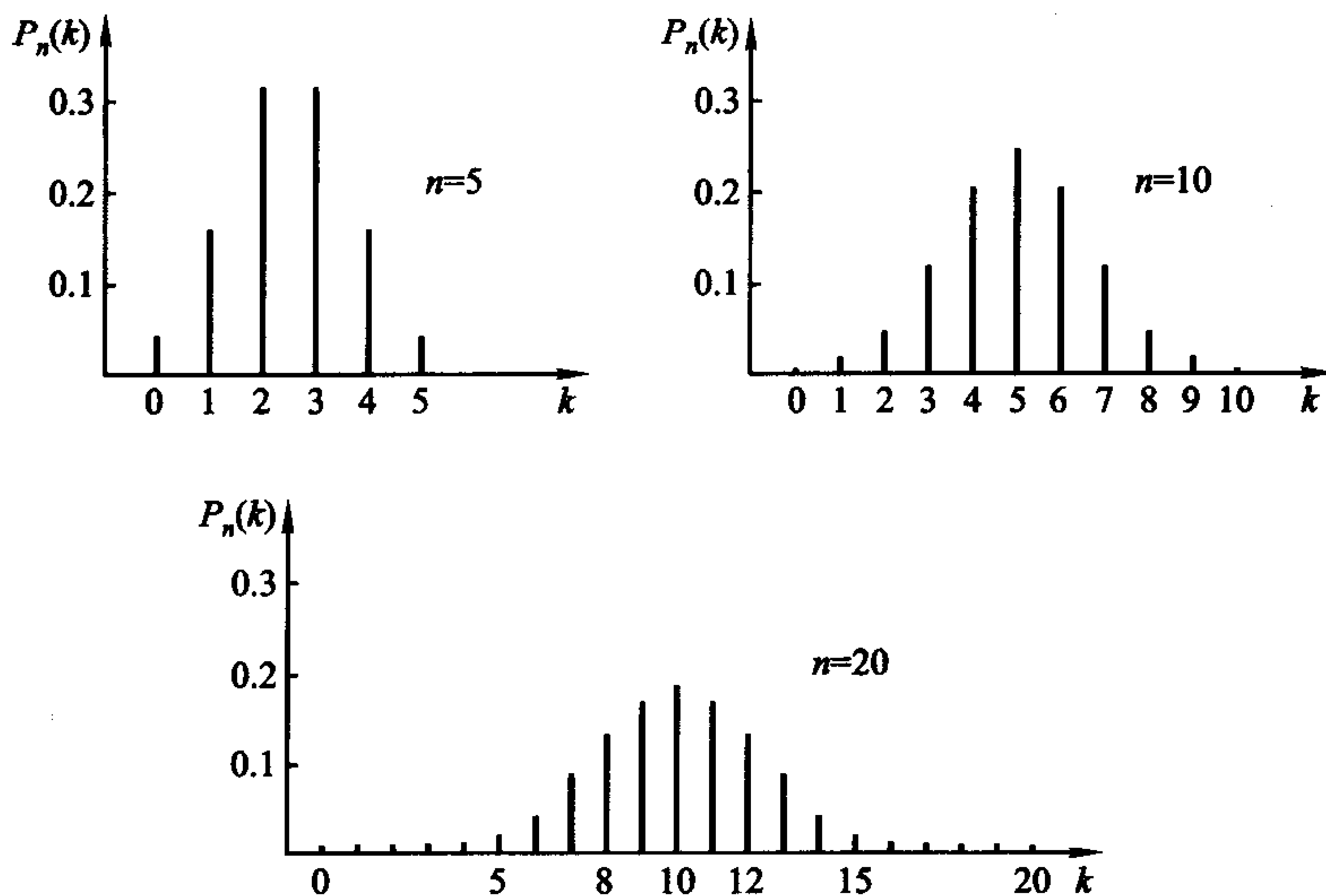


图 1 二项概率的示意图 $(n = 5, 10, 20)$

我们再引进一个 (实际上与上一个等价的) 模型, 它描绘某一质点的随机游动. 假设质点从 0 出发, 经过单位时间向上或向下移动一步 (图 2).

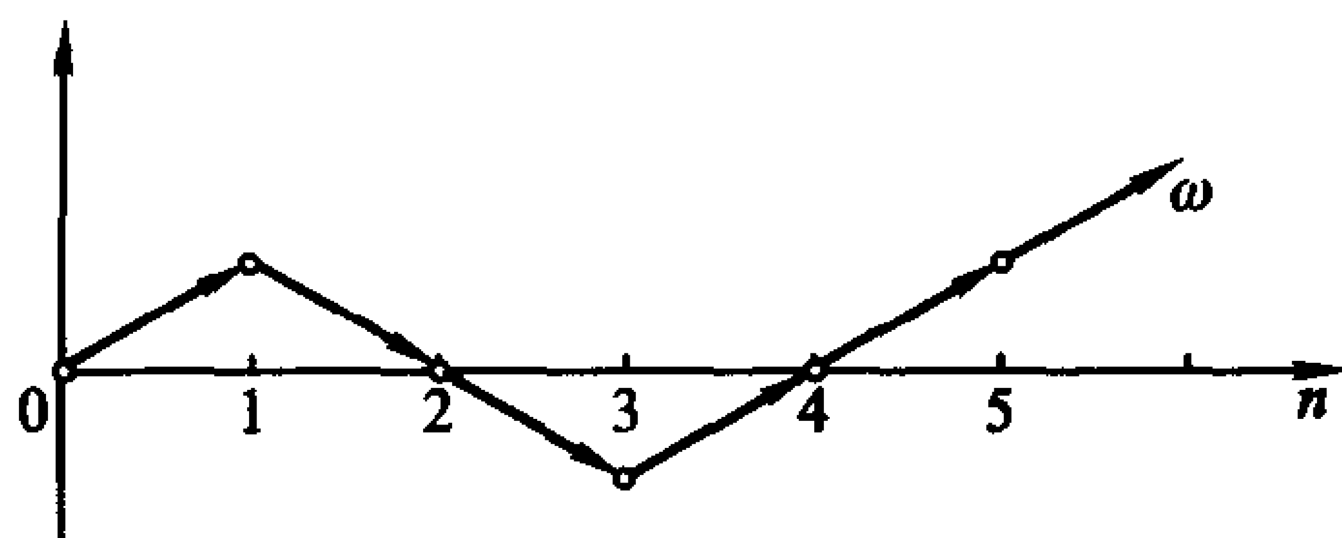


图 2

于是, 经过 n 步质点最多向上移动 n 步或向下移动 n 步. 显然, 质点运动的每一条 "轨道" ω, 可以完全由数组 $(a_1, \cdots, a_n)$ 描绘, 其中当质点在第 i 步向上移动时 $a_i = +1$, 而当质点在第 i 步向下移动时 $a_i = -1$. 赋予每条 "轨道" ω 以概率,

$$p(\omega) = p^{\nu(\omega)} q^{n-\nu(\omega)},$$

其中 $\nu(\omega)$ 是数组 $\omega = (a_1, \cdots, a_n)$ 中 "+1" 的个数, 即

$$\nu(\omega) = \frac{(a_1 + \cdots + a_n) + n}{2},$$

而 p 和 q 非负且 $p + q = 1$. 由于

$$\sum_{\omega \in \Omega} p(\omega) = 1,$$

可见概率组 $\{p(\omega)\}$, 连同轨道 $\omega = (a_1, \cdots, a_n)$ 的空间 Ω 及其子集, 确实决定质点 n 步运动的概率模型.

现在提出另一个问题: 事件 A_k "质点经 n 步到达纵坐标为 k 的点" 的概率如何? 一切满足 $\nu(\omega) - [n - \nu(\omega)] = k$, 即满足

$$\nu(\omega) = \frac{n+k}{2}$$

的轨道都满足上面提出的条件. 这样轨道的条数等于 $\mathrm{C}_n^{(n+k)/2}$ (见表 1 – 4), 因此

$$\mathbf{P}(A_k) = \mathrm{C}_n^{(n+k)/2} p^{(n+k)/2} q^{(n-k)/2}.$$

于是, 可以认为二项分布 $\{\mathbf{P}(A_{-n}), \cdots, \mathbf{P}(A_0), \cdots, \mathbf{P}(A_n)\}$, 描绘质点移动 n 步后位置的概率分布. 特别, 对于 "对称" (即 $p = q = 1/2$) 的情形, 每条轨道的概率等于 2^{-n}, 故

$$\mathbf{P}(A_k) = \mathrm{C}_n^{(n+k)/2} 2^{-n}.$$

下面讨论当 $n \to \infty$ 时这些概率的渐进性质.

如果移动步数等于 $2n$, 则由二项式系数的性质可见, 在概率 $\mathbf{P}(A_k)(|k| \leqslant 2n)$ 中最大概率为

$$\mathbf{P}(A_0) = \mathrm{C}_{2n}^{n} 2^{-2n}.$$

由斯特林 (J.Stirling) 公式*) (见 (6) 式)

$$n! \sim \sqrt{2\pi n}\mathrm{e}^{-n}n^n.$$

所以

$$\mathrm{C}_{2n}^n = \frac{(2n)!}{(n!)^2} \sim \frac{2^{2n}}{\sqrt{\pi n}},$$

即对充分大的 n,

$$\mathbf{P}(A_0) \sim \frac{1}{\sqrt{\pi n}}.$$

对于质点移动 $2n$ 步的情形, 图 3 是产生二项分布的示意图 (注意, 与图 2 不同, 时轴是铅直的, 并且正向朝上).

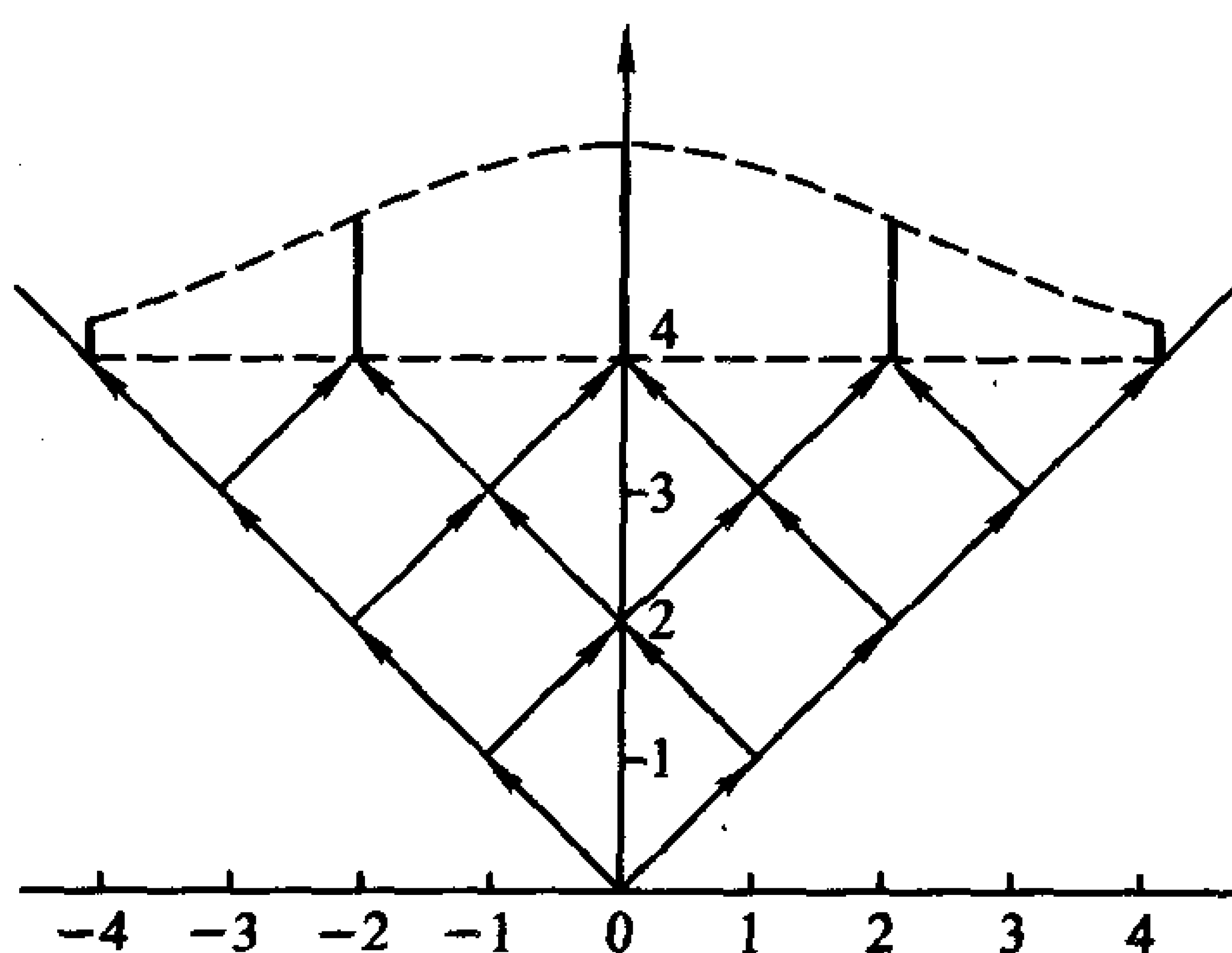


图 3 二项分布的产生

2. 多项分布 推广上述模型, 假设基本事件空间有如下构造

$$\Omega = \{\omega : \omega = (a_1, \cdots, a_n), a_i = b_1, \cdots, b_r\}$$

其中 $b_1, \cdots, b_r$ 是给定的数. 设 $\nu_i(\omega)$ 是序列 $\omega = (a_1, \cdots, a_n)$ 中等于 $b_i (i = 1, \cdots, r)$ 元素的个数, 而基本事件的 ω 概率等于

$$p(\omega) = p_1^{\nu_1(\omega)} \cdots p_r^{\nu_r(\omega)},$$

其中 $p_i \geqslant 0, p_1 + \cdots + p_r = 1$. 注意,

$$\sum_{\omega\in\Omega} p(\omega) = \sum_{\left\{\substack{n_1 \geqslant 0, \cdots, n_r \geqslant 0 \\ n_1 + \cdots + n_r = 1}\right\}} \mathrm{C}_n(n_1, \cdots, n_r) p_1^{n_1} \cdots p_r^{n_r},$$

*) 关系式 $f(n) \sim g(n)$ 表示 $f(n)/g(n) \to 1 (n \to \infty)$.

和式中以 $C_n(n_1,\cdots,n_r)$ 表示有序数组 $(a_1,\cdots,a_n)$ 的个数, 其中元素 b_1 重复 n_1 次,$\cdots$ 元素 b_r 重复 n_r 次. 因为, 有 $C_n^{n_1}$ 种方法将 n_1 个元素 b_1 安排在 n 个位置上, 有 $C_{n-n_1}^{n_2}$ 种方法将 n_2 个元素 b_2 安排在 $n-n_1$ 个位置上, 等等, 所以

$$\begin{aligned}C_n(n_1,\cdots,n_r) &= C_n^{n_1}C_{n-n_1}^{n_2}\cdots C_{n-(n_1+\cdots+n_{r-1})}^{n_r}\\ &= \frac{n!}{n_1!(n-n_1)!}\times\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\times\cdots\times 1=\frac{n!}{n_1!\cdots n_r!}.\end{aligned}$$

因此

$$\sum_{\omega\in\Omega}p(\omega)=\sum_{\left\{\begin{smallmatrix}n_1\geqslant 0,\cdots,n_r\geqslant 0\\ n_1+\cdots+n_r=1\end{smallmatrix}\right\}}\frac{n!}{n_1!\cdots n_r!}p_1^{n_1}\cdots p_r^{n_r}=(p_1+\cdots+p_r)^n=1,$$

于是, 多项分布确实是概率分布.

设

$$A_{n_1,\cdots,n_r}=\{\omega:\nu_1(\omega)=n_1,\cdots,\nu_r(\omega)=n_r\},$$

则

$$\mathbf{P}(A_{n_1,\cdots,n_r})=C_n(n_1,\cdots,n_r)p_1^{n_1}\cdots p_r^{n_r}. \tag{2}$$

概率组 $\{\mathbf{P}(A_{n_1,\cdots,n_r})\}$ 称做*多项分布*.

注意, 多项分布及其特例 —— 二项分布的产生与放回抽样相联系.

3. 多元超几何分布　出现在不放回抽样中.

作为例子, 假设一箱子中有编号为 $1,2,\cdots,M$ 的 M 个不同的球, 其中 M_1 个球具有颜色 $b_1,\cdots,M_r$ 个球具有颜色 b_r, 且 $M_1+\cdots+M_r=M$. 现在从箱中进行 $n(n<M)$ 次不放回抽样. 基本事件空间为

$$\Omega=\{\omega:\omega=(a_1,\cdots,a_n),a_k\neq a_l,k\neq l;a_i=1,\cdots,M\},$$

而 $N(\Omega)=(M)_n$. 假设基本事件是等可能的, 而 $B_{n_1,\cdots,n_r}$ 表示事件: "n_1 个球具有颜色 $b_1,\cdots,n_r$ 个球具有颜色 b_r", 且 $n_1+\cdots+n_r=n$. 求事件 $B_{n_1,\cdots,n_r}$ 的概率. 容易证明,

$$N(B_{n_1,\cdots,n_r})=C_n(n_1,\cdots,n_r)(M_1)_{n_1}\cdots(M_r)_{n_r},$$

因此

$$\mathbf{P}(B_{n_1,\cdots,n_r})=\frac{N(B_{n_1,\cdots,n_r})}{N(\Omega)}=\frac{C_{M_1}^{n_1}\cdots C_{M_r}^{n_r}}{C_M^n}. \tag{3}$$

概率组 $\{\mathbf{P}(B_{n_1,\cdots,n_r})\}$ 称做*多元超几何分布*. 当 $r=2$ 时此分布简称为*超几何分布*, 因为其母函数是超几何函数.

多元超几何分布的构造相当复杂. 因为概率

$$\mathbf{P}(B_{n_1,n_2})=\frac{C_{M_1}^{n_1}C_{M_2}^{n_2}}{C_M^n},n_1+n_2=n,M_1+M_2=M \tag{4}$$

包含 9 个阶乘数. 易见, 如果当 $M \to \infty, M_1 \to \infty$ 时, 且 $M_1/M \to p$, 从而 $M_2/M \to 1-p$, 则

$$\mathbf{P}(B_{n_1,n_2}) \to \mathrm{C}_{n_1+n_2}^{n_2} p^{n_1}(1-p)^{n_2}. \tag{5}$$

换句话说, 在上述条件下, 超几何分布逼近二项分布. 这直观上是明显的, 因为当 M 和 M_1 (有限) 较大时, 由不放回抽样得到几乎与放回抽样一样的结果.

例 利用公式 (4), 求体育抽彩中猜中 6 个 "幸运" 号码, 其实际意义如下.

假设有编号为 $1, 2, \cdots, 49$ 的 49 个球, 其中 6 个球是 "幸运的" (例如,6 个红球, 其余是白球). 随意从中不放回抽出 6 个球. 问抽出的 6 个球全部都是 "幸运的" 概率如何? 设 $M = 49, M_1 = 6, n_1 = 6, n_2 = 0$, 则我们感兴趣的事件为

$$B_{6,0} = \{6\text{ 个球都是 “幸运的”}\},$$

而根据公式 (4), 其概率为

$$\mathbf{P}(B_{6,0}) = \frac{1}{\mathrm{C}_{49}^{6}} \approx 7.2 \times 10^{-8}.$$

4. 斯特林公式 随 n 的增大, 阶乘数 $n!$ 增长的非常快. 例如,

$$10! = 3\ 628\ 800,$$

$$15! = 1\ 307\ 674\ 368\ 000,$$

而 100! 包含 158 位数字. 因此, 无论是从理论上还是实际计算的角度, **斯特林公式**

$$n! = \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n \exp\left\{\frac{\theta_n}{12n}\right\}, 0 < \theta_n < 1$$

都很重要. 数学分析的大量文献中, 都有斯特林公式证明 (亦可参见第八章 §8 练习题 1).

5. 练习题

1. 证明 (5) 式的命题.

2. 证明, 对于多项分布的概率, 在满足 $np_i - 1 < k_i \leqslant (n+r-1)p_i (i = 1, 2, \cdots, r)$ 的点 $(k_1, \cdots, k_r)$, 达到最大值.

3. **一维伊金格** (Ezenger) **模型**. 有 n 个质点, 分别位于编号 $1, 2, \cdots, n$ 的点. 假设质点分为两种类型: 类型 I 有 n_1 个质点, 类型 II 有 n_2 个质点, $n_1 + n_2 = n$, 并且全部 n 个质点的 $n!$ 排列是等可能的.

建立相应的概率模型, 并求事件

$$A_n(m_{11}, m_{12}, m_{21}, m_{22}) = \{\nu_{11} = m_{11}, \nu_{12} = m_{12}, \nu_{21} = m_{21}, \nu_{22} = m_{22}\}$$

的概率, 其中 ν_{ij} —— 排在 II 型质点 j 之后的 I 型质点 i 的个数 $(i, j = 1, 2)$.

4. 利用概率和组合的方法, 证明下列恒等式:

$$\sum_{k=0}^{n} \mathrm{C}_n^k = 2^n, \qquad \sum_{k=0}^{n} (\mathrm{C}_n^k)^2 = \mathrm{C}_{2n}^n,$$

$$\sum_{k=0}^{n} (-1)^{n-k} \mathrm{C}_m^k = \mathrm{C}_{m-1}^n, m \geqslant n+1,$$

$$\sum_{k=0}^{m} k(k-1)\mathrm{C}_m^k = m(m-1)2^{m-2}, m \geqslant 2,$$

$$k\mathrm{C}_n^k = n\mathrm{C}_{n-1}^{k-1},$$

$$\mathrm{C}_n^m = \sum_{j=0}^{m} \mathrm{C}_k^j \mathrm{C}_{n-k}^{m-j},$$

$(0 \leqslant m \leqslant n, 0 \leqslant k \leqslant n,$ 对于 $j < 0$ 和 $j > l$, 设 $\mathrm{C}_l^j = 0)$.

5. 假设 N 是某个总体的容量, 要求在对总体的全部元素没有简单重复计数的情况下, 以 "最少的费用" 估计 N. 例如, 在估计某个国家、城市 …… 人口时, 对类似的问题感兴趣.

拉普拉斯在 1786 年法国人口为 N 时, 曾经提出下面的方法.

从总体中选择若干个元素 (例如选择 M 个元素), 并且作上标记. 然后将这 M 个元素放回原总体, 并且与无标记的元素 "很好地混合". 然后从 "混合好了的" 总体中再抽取 n 个元素, 以 X 表示其中有标记的元素个数.

(1) 证明, 由超几何分布的公式 (4), 相应的概率 $\mathbf{P}_{N,M,n}\{X=m\}$ 可以表示为:

$$\mathbf{P}_{N,M,n}\{X=m\} = \frac{\mathrm{C}_M^n \mathrm{C}_{N-M}^{n-m}}{\mathrm{C}_N^n}.$$

(2) 假设 M, n 和 m 固定, 对 N 求上面概率的最大值, 即求总体的 "最大似然" 容量 N, 使 (对于给定的 M, n) 有标记的元素的个数 X 等于 m.

(3) 证明, 总体容量的最大似然估计值 (记作 $\widehat{N}$), 由公式表示,

$$\widehat{N} = [Mnm^{-1}],$$

其中 $[\,\cdot\,]$ 表示整数部分.

这样得到的估计量 $\widehat{N}$, 称做 N 的*最大似然估计量*. (在 §7 练习题 4 将继续讨论该问题).

6. (对照 §1 练习题 2.) 设 Ω 含 N 个元素, 而 $\widetilde{d}(N)$ 表示具有如下性质的不同分割的个数: 分割的每一个子集有奇数个元素. 证明

$$\widetilde{d}(1) = 1, \quad \widetilde{d}(2) = 1, \quad \widetilde{d}(3) = 2,$$
$$\widetilde{d}(4) = 5, \quad \widetilde{d}(6) = 12, \quad \widetilde{d}(7) = 37,$$

而一般, 有

$$\sum_{n=1}^{\infty}\frac{\widetilde{d}(n)x^n}{n!}=\mathrm{e}^{\operatorname{sh}x}-1,\quad |x|<1.$$

§3. 条件概率. 独立性

1. 事件的条件概率 事件的概率的概念, 可以回答下面一类问题: 假如箱中有 M 个球, 其中 M_1 个白球和 M_2 个黑球; 以 A 表示事件: 随意抽出一个球是白球. 问事件 A 的概率 $\mathbf{P}(A)$ 如何? 按古典型概率 $\mathbf{P}(A)=M_1/M$.

下面引进的条件概率的概念, 可以回答这样一类问题: 在第一次抽到白球 (事件 A) 的条件下, 在第二次也抽到白球 (事件 B) 的概率如何? (这里讨论的是不放回抽样).

这里自然应该这样讨论: 假如第一次抽到的球是白色的, 那么第二次抽球前箱中还有 $M-1$ 个球, 其中 M_1-1 个白球, 另外还有 M_2 个黑球; 直观上显然, 我们感兴趣的 (条件) 概率等于 $(M_1-1)/(M-1)$.

现在给出与直观概念一致的, 条件概率的定义.

设 $(\Omega,\mathscr{A},\mathbf{P})$ 是 (有限) 概率空间, 而 A 是某个事件 (即 $A\in\mathscr{A}$).

定义 1 设 $\mathbf{P}(A)>0$. 称

$$\frac{\mathbf{P}(AB)}{\mathbf{P}(A)} \tag{1}$$

为在事件 A 的条件下, 事件 B 的条件概率 (记作 $\mathbf{P}(B|A)$).

对于古典方法 (§1 的第 4 小节), 概率定义为

$$\mathbf{P}(A)=\frac{N(A)}{N(\Omega)},\quad \mathbf{P}(AB)=\frac{N(AB)}{N(\Omega)},$$

则

$$\mathbf{P}(B|A)=\frac{N(AB)}{N(A)}. \tag{2}$$

由定义 1 直接得出, 条件概率如下的性质:

$$\begin{aligned}
&\mathbf{P}(A|A)=1,\\
&\mathbf{P}(\varnothing|A)=0,\\
&\mathbf{P}(B|A)=1,\quad B\supseteq A,\\
&\mathbf{P}(B_1+B_2|A)=\mathbf{P}(B_1|A)+\mathbf{P}(B_2|A).
\end{aligned}$$

由这些性质可见, 对于固定的事件 A, 在概率空间 $(\Omega\bigcap A,\mathscr{A}\bigcap A)$ 上的条件概率 $\mathbf{P}(\cdot|A)$, 以及在空间 $(\Omega,\mathscr{A})$ 上的概率 $\mathbf{P}(\cdot)$ 具有同样的性质, 其中

$$\mathscr{A}\bigcap A=\{B\bigcap A: B\in\mathscr{A}\}.$$

注意,

$$\mathbf{P}(B|A)+\mathbf{P}(\overline{B}|A)=1,$$

然而, 一般

$$\mathbf{P}(B|A)+\mathbf{P}(B|\overline{A})\neq 1,$$
$$\mathbf{P}(B|A)+\mathbf{P}(\overline{B}|\overline{A})\neq 1.$$

例 1 考虑有两个孩子的家庭. 问一家中有两个男孩的概率如何? 假设

a) 岁数较大的是男孩;

b) 至少有一个是男孩.

显然, 基本事件空间为 (b —— 男孩, g —— 女孩):

$$\Omega=\{\mathrm{bb},\mathrm{bg},\mathrm{gb},\mathrm{gg}\},$$

其中 bg 表示 "岁数较大的是男孩, 岁数较小的是女孩", 等等.

假设 4 个基本事件都是等可能的. 那么,

$$\mathbf{P}(\mathrm{bb})=\mathbf{P}(\mathrm{bg})=\mathbf{P}(\mathrm{gb})=\mathbf{P}(\mathrm{gg})=\frac{1}{4}.$$

设事件 A"年长的是男孩", B"年幼的是男孩". 因此, $A\bigcup B$ 表示事件 "至少一个是男孩", 而 $A\bigcap B$ 表示事件 "两个全都是男孩". 我们感兴趣的问题是条件概率: a) $\mathbf{P}(AB|A)$, b) $\mathbf{P}(AB|A\bigcup B)$:

$$\mathbf{P}(AB|A)=\frac{\mathbf{P}(AB)}{\mathbf{P}(A)}=\frac{1/4}{1/2}=\frac{1}{2},$$
$$\mathbf{P}(AB|A\bigcup B)=\frac{\mathbf{P}(AB)}{\mathbf{P}(A\bigcup B)}=\frac{1/4}{3/4}=\frac{1}{3}.$$

2. 全概率公式和乘法公式 另一个简单而重要的公式 (3), 称做*全概率公式*, 是利用条件概率计算复合事件的基本工具.

考虑基本事件空间 Ω 的某个分割 $\mathscr{D}=\{A_1,\cdots,A_n\}$, 且 $\mathbf{P}(A_i)>0(i=1,2,\cdots,n)$. (这样的分割又称做不相容事件的*完全事件组*.) 显然,

$$B=BA_1+\cdots+BA_n,$$

因此

$$\mathbf{P}(B)=\sum_{i=1}^{n}\mathbf{P}(BA_i),$$

其中

$$\mathbf{P}(BA_i)=\mathbf{P}(B|A_i)\mathbf{P}(A_i).$$

从而, 有**全概率公式**:

$$\mathbf{P}(B)=\sum_{i=1}^{n}\mathbf{P}(B|A_i)\mathbf{P}(A_i). \tag{3}$$

特别, 如果 $0<\mathbf{P}(A)<1$, 则

$$\mathbf{P}(B)=\mathbf{P}(B|A)\mathbf{P}(A)+\mathbf{P}(B|\overline{A})\mathbf{P}(\overline{A}). \tag{4}$$

例 2 箱中有 M 个球, 其中 m 个是 "幸运的". 现在不放回地从中先后抽取两个球, 求后抽到的球是 "幸运的" 概率. 假设所有结局是等可能的, 而且关于先抽到的第一个球的情况未知. 设事件 $A=\{$ 先抽到的球是 "幸运的"$\}$, $B=\{$ 后抽到的球是 "幸运的"$\}$. 那么,

$$\mathbf{P}(B|A)=\frac{\mathbf{P}(AB)}{\mathbf{P}(A)}=\frac{\dfrac{m(m-1)}{M(M-1)}}{\dfrac{m}{M}}=\frac{m-1}{M-1},$$

$$\mathbf{P}(B|\overline{A})=\frac{\mathbf{P}(\overline{A}B)}{\mathbf{P}(\overline{A})}=\frac{\dfrac{m(M-m)}{M(M-1)}}{\dfrac{M-m}{M}}=\frac{m}{M-1},$$

从而, 由全概率公式, 有

$$\begin{aligned}\mathbf{P}(B)&=\mathbf{P}(B|A)\mathbf{P}(A)+\mathbf{P}(B|\overline{A})\mathbf{P}(\overline{A})\\&=\frac{m-1}{M-1}\times\frac{m}{M}+\frac{m}{M-1}\times\frac{M-m}{M}=\frac{m}{M}.\end{aligned}$$

有趣的是概率 $\mathbf{P}(A)$ 也等于 m/M. 这样, 尽管不知道先抽出的球的情况, 但是后抽出的是 "幸运球" 的概率并没有改变.

由条件概率的定义知, 当 $\mathbf{P}(A)>0$ 时, 有

$$\mathbf{P}(AB)=\mathbf{P}(B|A)\mathbf{P}(A). \tag{5}$$

该式称做**乘法公式**. 用数学归纳法可以将其推广: 假设事件组 $A_1,\cdots,A_n$ 满足条件: $\mathbf{P}(A_1\cdots A_n)>0$, 则

$$\mathbf{P}(A_1\cdots A_n)=\mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\cdots\mathbf{P}(A_n|A_1\cdots A_{n-1}), \tag{6}$$

其中 $A_1A_2\cdots A_n=A_1\bigcap A_2\bigcap\cdots\bigcap A_n$.

3. 贝叶斯公式 设事件 A 和 B 的概率大于 0: $\mathbf{P}(A)>0,\mathbf{P}(B)>0$, 则与 (5) 式同样, 有

$$\mathbf{P}(AB)=\mathbf{P}(A|B)\mathbf{P}(B). \tag{7}$$

由 (5) 和 (7) 式得所谓**贝叶斯** (Bayes) **公式**:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A)\mathbf{P}(B|A)}{\mathbf{P}(B)}. \tag{8}$$

假如事件组 $A_1, \cdots, A_n$ 是 Ω 的一个分割, 则由 (3) 和 (8) 式得

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B|A_i)}{\sum_{j=1}^{n} \mathbf{P}(A_j)\mathbf{P}(B|A_j)}, \tag{9}$$

称做**贝叶斯定理**.

在统计应用中, 事件 $A_1, \cdots, A_n$ 组成事件组 $(A_1 + \cdots + A_n = \Omega)$, 常称做 "假设" 或 "假说", 而 $\mathbf{P}(A_i)$ 称做假设 A_i 的验前概率*). 条件概率 $\mathbf{P}(A_i|B)$ 称做假设 A_i 在事件 B 出现后的验后概率.

例 3 假设匣中有两枚硬币: A_1 —— 对称的硬币, "正面" Z 出现的概率等于 1/2, 而 A_2 —— 不对称的硬币, "正面" Z 出现的概率等于 1/3. 随意选出一枚硬币并将其投掷, 结果掷出正面. 问抽到硬币为对称硬币的概率如何?

建立相应的概率模型. 这里自然取集合 $\Omega = \{A_1\mathrm{Z}, A_1\mathrm{F}, A_2\mathrm{Z}, A_2\mathrm{F}\}$, 可以描绘选取和投掷的结局, 其中 $A_1\mathrm{Z}$ 表示 "选中硬币" A_1, 结果掷出正面 Z, 等等, 而 F 表示硬币掷出反面. 根据条件, 所考虑结局的概率应该是:

$$\mathbf{P}(A_1) = \mathbf{P}(A_2) = \frac{1}{2}$$

和

$$\mathbf{P}(\mathrm{Z}|A_1) = \frac{1}{2}, \quad \mathbf{P}(\mathrm{Z}|A_2) = \frac{1}{3}.$$

这些条件唯一决定各结局的概率:

$$\mathbf{P}(A_1\mathrm{Z}) = \frac{1}{4}, \mathbf{P}(A_1\mathrm{F}) = \frac{1}{4}, \mathbf{P}(A_2\mathrm{Z}) = \frac{1}{6}, \mathbf{P}(A_2\mathrm{F}) = \frac{1}{3}.$$

那么, 根据贝叶斯公式, 所求的概率为

$$\mathbf{P}(A_1|\mathrm{Z}) = \frac{\mathbf{P}(A_1)\mathbf{P}(\mathrm{Z}|A_1)}{\mathbf{P}(A_1)\mathbf{P}(\mathrm{Z}|A_1) + \mathbf{P}(A_2)\mathbf{P}(\mathrm{Z}|A_2)} = \frac{3}{5},$$

从而

$$\mathbf{P}(A_2|\mathrm{Z}) = \frac{2}{5}.$$

---

*) a priori—— 验前, a posteriori—— 验后.

4. 独立性 在这一小节将引进的独立性的概念, 它对于概率论在一定意义上有核心作用: 正是独立性的概念决定了概率论的特色, 使它从研究有测度的可测空间的一般理论中分离出来.

对于两个事件 A 和 B, 如果事件 A 的出现, 对事件 B 出现的概率不产生丝毫影响, 自然应该认为事件 B 不依赖于事件 A. 换句话说, 称 "事件 B 对事件 A 独立", 如果

$$\mathbf{P}(B|A) = \mathbf{P}(B), \tag{10}$$

其中假设 $\mathbf{P}(A) > 0$.

因为

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)},$$

所以由 (10) 可见

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B). \tag{11}$$

同样, 设 $\mathbf{P}(B) > 0$, 则自然称 "事件 A 对事件 B 独立", 如果

$$\mathbf{P}(A|B) = \mathbf{P}(A).$$

由此仍然得到关系式 (11), 于是该式关于事件 A 和 B 是对称的, 并且当事件 A 和 B 的概率 (一个或两个) 可能为 0 时 (11) 式仍然成立. 由此导出独立性的定义.

定义 2 称事件 A 和 B (关于概率 $\mathbf{P}$) 为**独立的**或**统计独立的**, 如果

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B).$$

在概率论中, 往往不但需要考虑事件 (集合) 的独立性, 而且需要研究事件 (集合) 组的独立性. 下面引进相应的定义.

定义 3 称 Ω 子集系的代数 $\mathscr{A}_1$ 和 $\mathscr{A}_2$ (关于概率 $\mathbf{P}$) 为**独立的**或**统计独立的**, 如果对于相应地属于 $\mathscr{A}_1$ 和 $\mathscr{A}_2$ 的两个任意子集 A_1 和 A_2 独立.

作为例子, 考虑两个代数:

$$\mathscr{A}_1 = \{A_1, \overline{A}_1, \varnothing, \Omega\} \quad 和 \quad \mathscr{A}_2 = \{A_2, \overline{A}_2, \varnothing, \Omega\}$$

其中 A_1 和 A_2 分别是 $\mathscr{A}_1$ 和 $\mathscr{A}_2$ 中的集合. 不难证明, 两个集合代表 $\mathscr{A}_1$ 和 $\mathscr{A}_2$ 独立当且仅当对任意事件 A_1 和 A_2 独立. 事实上, $\mathscr{A}_1$ 和 $\mathscr{A}_2$ 独立说明, 16 个事件偶 A_1 和 A_2, A_1 和 $\overline{A}_2, \cdots, \Omega$ 和 Ω 独立. 从而 A_1 和 A_2 独立. 相反, 若 A_1 和 A_2 独立, 则需要证明其余 15 个事件偶也独立. 例如, 现在验证 A_1 和 $\overline{A}_2$ 独立. 有

$$\begin{aligned}\mathbf{P}(A_1\overline{A}_2) &= \mathbf{P}(A_1) - \mathbf{P}(A_1A_2) = \mathbf{P}(A_1) - \mathbf{P}(A_1)\mathbf{P}(A_2) \\ &= \mathbf{P}(A_1)[1 - \mathbf{P}(A_2)] = \mathbf{P}(A_1)\mathbf{P}(\overline{A}_2).\end{aligned}$$

同样可以证明其余事件偶独立.

5. 全体独立和两两独立 两个集合以及两个集合代数独立的概念, 可以推广到任意有限个集合以及集合代数的情形.

定义 4 称集合 $A_1, \cdots, A_n$ (关于概率 $\mathbf{P}$) *全体独立*或*全体统计独立*, 如果对于任何 $k = 1, 2, \cdots, n$ 和 $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$,

$$\mathbf{P}(A_{i_1}, \cdots, A_{i_k}) = \mathbf{P}(A_{i_1}) \cdots \mathbf{P}(A_{i_k}). \tag{12}$$

定义 5 称集合 (关于概率 $\mathbf{P}$) 代数 $\mathscr{A}_1, \cdots, \mathscr{A}_n$ *全体独立*或*全体统计独立*, 如果相应地属于 $\mathscr{A}_1, \cdots, \mathscr{A}_n$ 的任何集合 $A_1, \cdots, A_n$ 独立.

需要指出, 由事件*两两独立*, 一般它们*未必*全体独立. 事实上, 例如, 若 $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ 且所有基本事件都等可能, 则事件

$$A = \{\omega_1, \omega_2\}, \quad B = \{\omega_1, \omega_3\}, \quad C = \{\omega_1, \omega_4\}$$

两两独立, 但

$$\mathbf{P}(ABC) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3 = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C).$$

还需要指出, 对于某些事件 A, B, C, 由

$$\mathbf{P}(ABC) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C),$$

一般并不能得出这些事件*两两独立*. 事实上, 设空间 Ω 由有序数偶组成: (i, j), 其中 $i, j = 1, 2, \cdots, 6$, 且这些数偶都等可能. 那么, 对于事件

$$A = \{(i,j) : j = 1, 2 \text{ 或 } 5\}, \quad B = \{(i,j) : j = 4, 5 \text{ 或 } 6\}, \quad C = \{(i,j) : i + j = 9\},$$

则

$$\begin{aligned}
\mathbf{P}(AB) &= \frac{1}{6} \neq \frac{1}{4} = \mathbf{P}(A)\mathbf{P}(B), \\
\mathbf{P}(AC) &= \frac{1}{36} \neq \frac{1}{18} = \mathbf{P}(A)\mathbf{P}(C), \\
\mathbf{P}(BC) &= \frac{1}{12} \neq \frac{1}{18} = \mathbf{P}(B)\mathbf{P}(C),
\end{aligned}$$

然而

$$\mathbf{P}(ABC) = \frac{1}{36} = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C).$$

6. 伯努利概型 从*独立性*的概念出发, 详细讨论 §2 中引进的导出二项分布的经典模型 $(\Omega, \mathscr{A}, \mathbf{P})$. 在此模型中,

$$\begin{aligned}
\Omega &= \{\omega : \omega = (a_1, \cdots, a_n), a_i = 0, 1\}, \\
\mathscr{A} &= \{A : A \subseteq \Omega\},
\end{aligned}$$

而 $\mathbf{P}(\{\omega\}) = p(\omega)$, 其中

$$p(\omega) = p^{\sum a_i}(1-p)^{n-\sum a_i}. \tag{13}$$

考虑事件 $A \subseteq \Omega$. 如果事件 A 只决定于 a_k 的值, 则称该事件依赖于 k 时的试验. 事件 A_k 和 $\overline{A}_k$ 就是这样事件的例子:

$$A_k = \{\omega : a_k = 1\}, \overline{A}_k = \{\omega : a_k = 0\}.$$

考虑代数序列: $\mathscr{A}_1, \cdots, \mathscr{A}_n$, 其中 $\mathscr{A}_k = \{A_k, \overline{A}_k, \varnothing, \Omega\}$, 并证明对于 (13) 的情形, 这些代数独立.

显然,

$$\mathbf{P}(A_k) = \sum_{\{\omega:a_k=1\}} p(\omega) = \sum_{\{\omega:a_k=1\}} p^{\sum a_i}(1-p)^{n-\sum a_i}$$

$$= p \sum_{\{a_i:i\neq k\}} p^{\sum\limits_{i\neq k} a_i} q^{(n-1)-\sum\limits_{i\neq k} a_i} = p\sum_{l=1}^{n-1} \mathrm{C}_{n-1}^{l} p^l q^{(n-1)-l} = p,$$

而类似的计算, 可得 $\mathbf{P}(\overline{A}_k) = q$, 且当 $k \neq l$ 时, 有

$$\mathbf{P}(A_k A_l) = p^2, \mathbf{P}(A_k \overline{A}_l) = pq, \mathbf{P}(\overline{A}_k \overline{A}_l) = q^2.$$

由此容易证明, 代数 $\mathscr{A}_k$ 和 $\mathscr{A}_l (k \neq l)$ 独立.

类似地可以证明代数 $\mathscr{A}_1, \mathscr{A}_2, \cdots, \mathscr{A}_n$ 独立. 因此可以把所讨论的模型 $(\Omega, \mathscr{A}, \mathbf{P})$, 称做适合*具有两种结局且成功的概率为 p 的 n 次独立试验的模型*. 伯努利 (J. Bernoulli) 最早研究了该模型, 并且对这种模型证明了*大数定律* (§5). 因此, 该模型又称做 (具有 "成功" 与 "失败" 两种结局且 "成功" 概率为 p 的) *伯努利概型*.

深入研究伯努利概型的概率空间表明, 它具有下面将阐述的 "概率空间直积" 的构造.

假设有 n 个有限概率空间 $(\Omega_1, \mathscr{B}_1, \mathbf{P}_1), \cdots, (\Omega_n, \mathscr{B}_n, \mathbf{P}_n)$. 组成点 $\omega = (a_1, \cdots, a_n)$ 的空间 $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$, 其中 $a_i \in \Omega_i$. 记

$$\mathscr{A} = \mathscr{B}_1 \otimes \mathscr{B}_2 \otimes \cdots \otimes \mathscr{B}_n$$

是 Ω 的子集的代数, 由形如 $A = B_1 \times B_2 \times \cdots \times B_n (B_i \in \mathscr{B}_i)$ 的集合构成. 最后, 对于 $\omega = (a_1, \cdots, a_n)$, 令 $p(\omega) = p_1(a_1) \cdots p_n(a_n)$, 且把集合 $A = B_1 \times B_2 \times \cdots \times B_n$ 的概率定义为:

$$\mathbf{P}(A) = \sum_{\{a_1 \in B_1, \cdots, a_n \in B_n\}} p_1(a_1) \cdots p_n(a_n).$$

不难验证 $\mathbf{P}(\Omega) = 1$, 因此 $(\Omega, \mathscr{A}, \mathbf{P})$ 决定某个概率空间, 称做*概率空间*

$$(\Omega_1, \mathscr{B}_1, \mathbf{P}_1), \cdots, (\Omega_n, \mathscr{B}_n, \mathbf{P}_n)$$

的直积.

概率空间直积一条容易验证的性质是: 事件

$$A_1 = \{\omega : a_1 \in B_1\}, \cdots, A_n = \{\omega : a_n \in B_n\}$$

关于概率 $\mathbf{P}$ 独立, 其中 $B_i \in \mathscr{B}_i$. 同样, 空间 Ω 的子集代数

$$\mathscr{A}_1 = \{A_1 : A_1 = \{\omega : a_1 \in B_1\}, B_1 \in \mathscr{B}_1\},$$
$$\cdots\cdots\cdots\cdots$$
$$\mathscr{A}_n = \{A_n : A_n = \{\omega : a_n \in B_n\}, B_n \in \mathscr{B}_n\},$$

独立.

由上面引进的构造, 可见伯努利概型:

$$(\Omega, \mathscr{A}, \mathbf{P}), \Omega = \{\omega : \omega = (a_1, \cdots, a_n), a_i = 0, 1\},$$
$$\mathscr{A} = \{A : A \subseteq \Omega\} \quad \text{和} \quad \mathbf{P}(\{\omega\}) = p^{\sum a_i}(1-p)^{n-\sum a_i},$$

可以由概率空间 $(\Omega_1, \mathscr{B}_1, \mathbf{P}_1), \cdots, (\Omega_n, \mathscr{B}_n, \mathbf{P}_n)$ 的直积得到, 其中

$$\Omega_i = \{0, 1\}, \quad \mathscr{B}_i = \{\{0\}, \{1\}, \varnothing, \Omega_i\},$$
$$\mathbf{P}_i(\{1\}) = p, \quad \mathbf{P}_i(\{0\}) = q.$$

7. 练习题

1. 举例说明等式

$$\mathbf{P}(B|A) + \mathbf{P}(B|\overline{A}) = 1,$$
$$\mathbf{P}(B|A) + \mathbf{P}(\overline{B}|\overline{A}) = 1$$

一般不正确.

2. 箱中有 M 球, 其中 M_1 个白球. 考虑不放回抽取的容量为 n 的样本. 以 B_j 表示事件 "第 j 次抽到的是白球", 以 A_k 表示事件 "在容量为 n 的样本中恰好 k 个白球". 证明, 无论对于不放回抽样, 还是放回抽样, 都有

$$\mathbf{P}(B_j|A_k) = \frac{k}{n}.$$

3. 对于独立事件 $A_1, \cdots, A_n$, 证明

$$\mathbf{P}\left(\bigcup_{k=1}^{n} A_k\right) = 1 - \prod_{k=1}^{n} \mathbf{P}(\overline{A}_k).$$

4. 对于独立事件 $A_1, \cdots, A_n$, 其中 $\mathbf{P}(A_i) = p_i$. 证明这些事件一个都不出现的概率

$$p_0 = \prod_{i=1}^{n}(1 - p_i).$$

5. 设 A 和 B 是独立事件. 通过 $\mathbf{P}(A)$ 和 $\mathbf{P}(B)$ 表示下列事件的概率: 在 A 和 B 之中, 恰好出现 k 个事件, 至少出现 k 个事件, 以及最多出现 k 个事件.

6. 设事件 A 与它自己独立, 即 A 与 A独立, 证明 $\mathbf{P}(A)$ 等于 0 或 1.

7. 设事件 A 概率 $\mathbf{P}(A)$ 等于 0 或 1, 证明 A 与任何事件 B 独立.

8. 考虑图 4 的电路图. 在 A, B, C, D 和 E 等 5 个继电器中, 各继电器独立的工作, 每个以概率 p 和 q 分别有 "断" (无信号通过) 和 "通"(有信号通过) 两种状态. 问 "在入口发送的信号, 出口可以收到" 信号的概率如何? 在 E "处于 '通' 的状态下, 出口可以收到" 信号的概率如何?

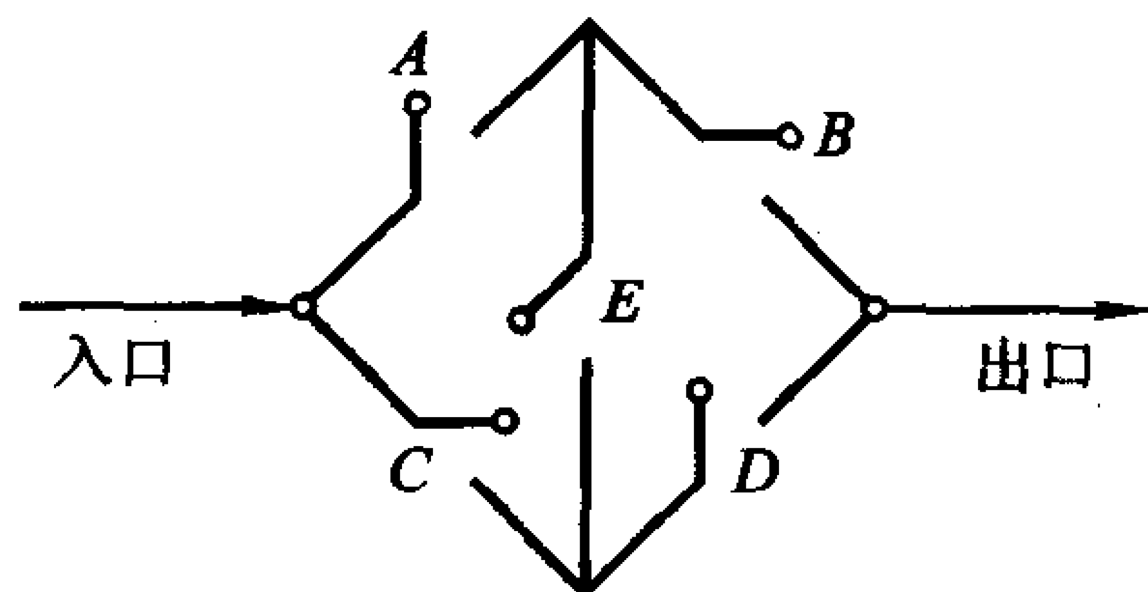


图 4

9. 设 $\mathbf{P}(A+B)>0$, 证明

$$\mathbf{P}(A|A+B)=\frac{\mathbf{P}(A)}{\mathbf{P}(A)+\mathbf{P}(B)}.$$

10. 设事件 A 对事件 $B_n (n \geqslant 1)$ 独立, 并且 $B_i \bigcap B_j=\varnothing, i \neq j$. 证明事件

$$A \quad 与 \quad \bigcup_{n=1}^{\infty} B_n$$

独立.

11. 证明, 如果 $\mathbf{P}(A|C)>\mathbf{P}(B|C)$ 和 $\mathbf{P}(A|\overline{C})>\mathbf{P}(B|\overline{C})$, 则 $\mathbf{P}(A)>\mathbf{P}(B)$.

12. 证明

$$\mathbf{P}(A|B)=\mathbf{P}(A|BC)\mathbf{P}(C|B)+\mathbf{P}(A|B\overline{C})\mathbf{P}(\overline{C}|B).$$

13. 设 X 和 Y 独立, 服从参数为 n 和 p 的二项分布, 证明

$$\mathbf{P}\{X=k|X+Y=m\}=\frac{\mathrm{C}_n^k \mathrm{C}_n^{m-k}}{\mathrm{C}_{2n}^m}, \quad k=0,1,\cdots,\min(m,n).$$

14. 设 A, B, C 是两两独立的等概率事件, 且 $A \bigcap B \bigcap C=\varnothing$. 求概率 $\mathbf{P}(A)$ 的最大值.

15. 在箱中原来有一个白球, 以相同的概率将一个白球或黑球放入箱中. 然后随意取出一个球, 结果是白球. 问箱中剩下的球也是白球的概率如何?

§4. 随机变量及其特征

1. 随机变量及其概率分布和分布函数 设 $(\Omega, \mathscr{A}, \mathbf{P})$ 是某具有有限个结局试验的概率模型, $N(\Omega)$ 是 Ω 中基本事件的个数, 而 $\mathscr{A}$ 是 Ω 中所有子集的代数. 能够理解, 以上讨论的各种事件 $A \in \mathscr{A}$ 概率的计算, 其实基本事件空间的自然本性并不重要. 重要的只是某种数字特征, 其值依赖于基本事件. 我们想知道的是, 在一系

列 n 次试验中, 出现一定次数成功的概率如何, 分到各箱中质点个数服从哪种概率分布, 等等.

现在要引进的随机变量的概念, 可以表征在随机试验中 "测量" 结果的量. 下面, 将再引进随机变量更一般的形式.

定义 1 称任一定义在 (有限) 基本事件空间 Ω 上的数值函数 $\xi = \xi(\omega)$ 为**随机变量**. (第二章 §4, 引进随机变量的一般概念之后, 就可以清楚知道随机变量的 "简单" 术语的来源.)

例 1 对于接连两次掷硬币模型, 其基本事件空间为 $\Omega = \{ZZ, ZF, FZ, FF\}$, 其中 Z —— 正面, F —— 反面. 我们利用下面的表格定义随机变量 $\xi = \xi(\omega)$:

ω	ZZ	ZF	FZ	FF
$\xi(\omega)$	2	1	1	0

这里, 按实际含义, $\xi(\omega)$ 是对应于 ω 的 "正面" 出现的次数.

随机变量 ξ 的另一简单的例子是某集合 $A \in \mathscr{A}$ 的**示性函数** (亦称**特征函数**):

$$\xi = I_A(\omega),$$

其中*)

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

当实验者遇到描绘某些记载或读数的随机变量时, 则他关心的基本问题是, 该随机变量取各个数值的概率如何. 从这种观点出发, 关心的不是概率 $\mathbf{P}$ 在 $(\Omega, \mathscr{A})$ 上的分布, 而是概率在随机变量之可能值的集合上的分布. 由于对于所研究的情形, Ω 由有限个点构成, 则随机变量 ξ 的值域 X 也是有限的. 设 $X = \{x_1, \cdots, x_m\}$, 其中 $x_1, \cdots, x_m$ 是 ξ 的全部可能值.

记 $\mathscr{X}$ 为值域 X 上一切子集的全体, 并设 $B \in \mathscr{X}$. 当 X 是随机变量 ξ 的值域时, 集合 B 也可以视为某个事件.

在 $(X, \mathscr{X})$ 上考虑由随机变量 ξ 按公式,

$$P_\xi(B) = \mathbf{P}\{\omega : \xi(\omega) \in B\}, \quad B \in \mathscr{X}$$

产生的概率 $P_\xi(\cdot)$. 显然, 这些概率的值完全决定于:

$$P_\xi(x_i) = \mathbf{P}\{\omega : \xi(\omega) = x_i\}, \quad x_i \in X.$$

组数 $\{P_\xi(x_1), \cdots, P_\xi(x_m)\}$ 称做随机变量 ξ 的**概率分布**.

---

*) 对于示性函数 $I_A(\omega)$, 还使用记号 $I(A), I_A$. 关于以后常用到的一些性质参见练习题 1.

例 2 设随机变量 ξ 分别以概率 p 和 q 取 1 和 0 两个值, 其中 p 称做 "成功" 的概率, 而 q 称做 "失败" 的概率, 则称 ξ 为*伯努利随机变量*[*]. 显然, 对于随机变量 ξ, 有

$$P_\xi(x) = p^x q^{1-x}, \quad x = 0, 1. \tag{1}$$

设 ξ 是以概率

$$P_\xi(x) = \mathrm{C}_n^x p^x q^{n-x}, \quad x = 0, 1, \cdots, n \tag{2}$$

取 $0, 1, \cdots, n$ 等 $n+1$ 个可能值, 则称 ξ 为*二项随机变量*, 或称 ξ 为*服从二项分布的随机变量*.

注意, 在这些及以后举的许多例子中, 我们不具体说明基本概率空间 $(\Omega, \mathscr{A}, \mathbf{P})$ 的构造, 而只关心随机变量的值及其概率分布.

随机变量 ξ 的构造完全由概率分布 $\{P_\xi(x_i), 1, 2, \cdots, m\}$ 描述. 下面引进的分布函数的概念, 提供随机变量构造的等价描述.

定义 2 *设 $x \in \mathbb{R}^1$. 函数*

$$F_\xi(x) = \mathbf{P}\{\omega : \xi(\omega) \leqslant x\}$$

称做随机变量 ξ 的*分布函数*,

显然,

$$F_\xi(x) = \sum_{\{i: x_i \leqslant x\}} p_\xi(x_i)$$

并且

$$P_\xi(x_i) = F_\xi(x_i) - F_\xi(x_i-),$$

其中

$$F_\xi(x-) = \lim_{y \uparrow x} F_\xi(y).$$

如果假设 $x_1 < x_2 < \cdots < x_m$, 而 $F_\xi(x_0) = 0$, 则

$$P_\xi(x_i) = F_\xi(x_i) - F_\xi(x_{i-1}), \quad i = 1, \cdots, m.$$

下面的图 5 是二项随机变量 ξ 的 $P_\xi(x)$ 和 $F_\xi(x)$ 的示意图.

直接由定义 2 可见, 分布函数具有下列性质:

(1) $F_\xi(-\infty) = 0, F_\xi(+\infty) = 1$;

(2) $F_\xi(x)$ *右连续*: $F_\xi(x+) = F_\xi(x)$, *并且是阶梯函数*.

---

[*] 在概率论的文献里, 所说的 "伯努利"、"二项"、"泊松"、"高斯" ······ 随机变量, 通常称做随机变量服从伯努利、二项、泊松、高斯 ······ 分布.

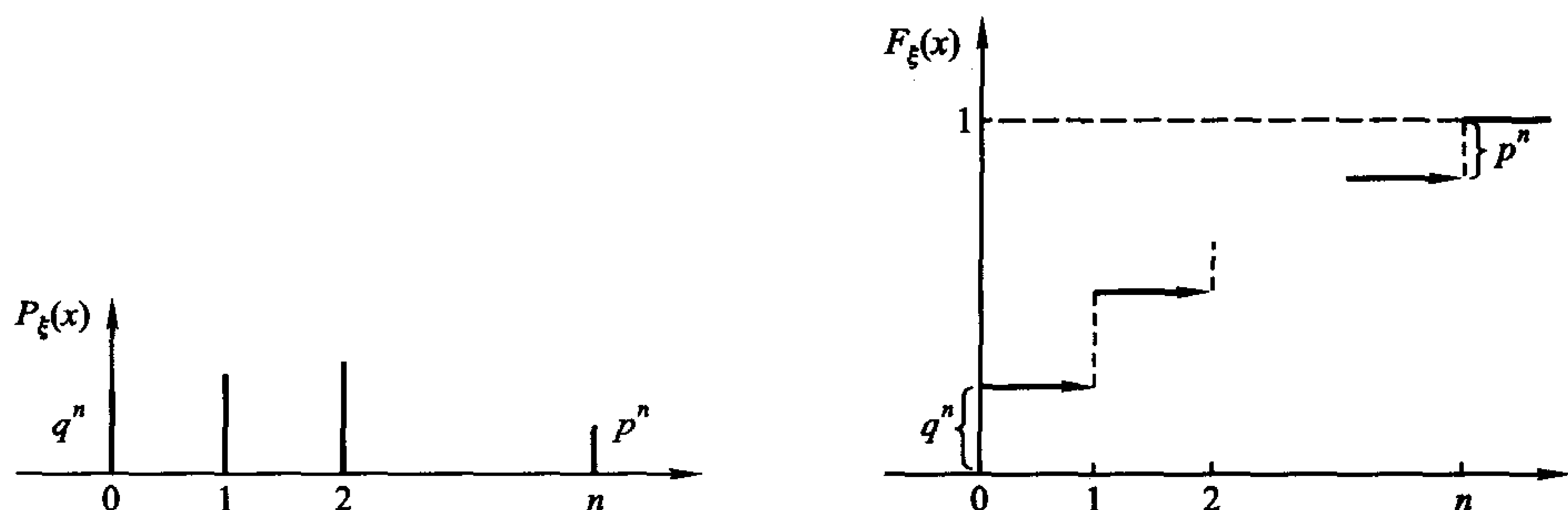


图 5

除随机变量之外, 还常研究随机向量 $\xi=(\xi_1,\cdots,\xi_r)$, 其各分量都是随机变量. 例如, 在研究多项分布时, 对象就是随机向量 $\nu=(\nu_1,\cdots,\nu_r)$, 其中 $\nu_i=\nu_i(\omega)$ 是序列 $\omega=(a_1,\cdots,a_n)$ 中等于 $b_i(i=1,\cdots,r)$ 的分量的个数.

对于 $x_i\in X_i(X_i(i=1,\cdots,r)$ 是 ξ_i 的一切可能值的集合), 概率

$$P_\xi(x_1,\cdots,x_r)=\mathbf{P}\{\omega:\xi_1(\omega)=x_1,\cdots,\xi_r(\omega)=x_r\}$$

的全体, 称做随机向量 $\xi=(\xi_1,\cdots,\xi_r)$ 的概率分布, 而函数

$$F_\xi(x_1,\cdots,x_r)=\mathbf{P}\{\omega:\xi_1(\omega)\leqslant x_1,\cdots,\xi_r(\omega)\leqslant x_r\}$$

称做随机向量 $\xi=(\xi_1,\cdots,\xi_r)$ 的分布函数, 其中 $x_i\in\mathbb{R}^1(i=1,\cdots,r)$.

对于上面提到的向量 $\nu=(\nu_1,\cdots,\nu_r)$,

$$P_\nu(n_1,\cdots,n_r)=\mathrm{C}_n(n_1,\cdots,n_r)p_1^{n_1}\cdots p_r^{n_r}$$

(见 §2 中的 (2) 式).

2. 独立随机变量 设是一组 $\xi_1,\cdots,\xi_r$ 在 $\mathbb{R}^1$ 中 (有限) 集合 X 上取值的随机变量. 记 $\mathscr{X}$ 是 X 中所有子集的代数.

定义 3 称随机变量 $\xi_1,\cdots,\xi_r$ 为 (全体) 独立的, 如果对于任意 $x_1,\cdots,x_r\in X$,

$$\mathbf{P}\{\xi_1=x_1,\cdots,\xi_r=x_r\}=\mathbf{P}\{\xi_1=x_1\}\cdots\mathbf{P}\{\xi_r=x_r\},$$

或等价地: 对于任意 $B_1,\cdots,B_r\in\mathscr{X}$,

$$\mathbf{P}\{\xi_1\in B_1,\cdots,\xi_r\in B_r\}=\mathbf{P}\{\xi_1\in B_1\}\cdots\mathbf{P}\{\xi_r\in B_r\}.$$

上面讨论的伯努利概型, 就是独立随机变量的一个最简单例子. 具体地, 设

$$\Omega=\{\omega:\omega=(a_1,\cdots,a_n),a_i=0,1\},\quad p(\omega)=p^{\sum a_i}(1-p)^{n-\sum a_i},$$

并且对于 $\omega=(a_1,\cdots,a_n)$, $\xi_i(\omega)=a_i(i=1,\cdots,n)$. 那么, 由 §3 的证明, 事件

$$A_1=\{\omega:a_1=1\},\cdots,A_n=\{\omega:a_n=1\}$$

独立, 从而 $\xi_1,\cdots,\xi_n$ 独立.

3. 随机变量之和的概率分布 我们以后不止一次地遇到，作为随机变量 $\xi_1, \cdots, \xi_r$ 的函数的随机变量 $f(\xi_1, \cdots, \xi_r)$ 的概率分布问题. 现在只限于考虑求随机变量之和 $\xi + \eta$ 的分布的情形.

如果 ξ 的值域为 $X = \{x_1, \cdots, x_k\}$, η 的值域为 $Y = \{y_1, \cdots, y_l\}$, 则 $\zeta = \xi + \eta$ 的值域为 $Z = \{z : z = x_i + y_j, i = 1, \cdots, k, j = 1, \cdots, l\}$, 且显然

$$P_\zeta(z) = \mathbf{P}\{\zeta = z\} = \mathbf{P}\{\xi + \eta = z\}$$
$$= \sum_{\{(i,j):x_i+y_j=z\}} \mathbf{P}\{\xi = x_i, \eta = y_j\}.$$

随机变量 ξ 和 η 独立的情形特别重要, 即 $\mathbf{P}\{\xi = x_i, \eta = y_j\} = \mathbf{P}\{\xi = x_i\}\mathbf{P}\{\eta = y_j\}$. 这时, 对于任意 $z \in Z$

$$P_\zeta(z) = \sum_{\{(i,j):x_i+y_j=z\}} P_\xi(x_i)P_\eta(y_j) = \sum_{i=1}^{k} P_\xi(x_i)P_\eta(z - x_i). \tag{3}$$

例如, 若 ξ 和 η 是分别以概率 p 和 q 取 1 和 0 为值的独立伯努利随机变量, 则 $Z = \{0, 1, 2\}$, 而

$$P_\zeta(0) = P_\xi(0)P_\eta(0) = q^2,$$
$$P_\zeta(1) = P_\xi(0)P_\eta(1) + P_\xi(1)P_\eta(0) = 2pq,$$
$$P_\zeta(2) = P_\xi(1)P_\eta(1) = p^2.$$

用归纳法容易证明, 若 $\xi_1, \cdots, \xi_r$ 是独立伯努利随机变量, 且 $\mathbf{P}\{\xi_i = 1\} = p, \mathbf{P}\{\xi_i = 0\} = q$, 则随机变量 $\zeta = \xi_1 + \cdots + \xi_r$ 服从二项分布

$$P_\zeta(k) = \mathrm{C}_n^k p^k q^{n-k} \quad (k = 0, 1, \cdots, n). \tag{4}$$

4. 数学期望 现在研究重要概念, 随机变量的数学期望或均值.

设 $(\Omega, \mathscr{A}, \mathbf{P})$ 是 (有限) 概率空间, 而 $\xi = \xi(\omega)$ 是某一随机变量, 其值域为 $X = \{x_1, \cdots, x_k\}$. 如果设 $A_i = \{\omega : \xi(\omega) = x_i\}$, 则显然 ξ 可以表示为

$$\xi(\omega) = \sum_{i=1}^{k} x_i I(A_i), \tag{5}$$

其中集合 $A_1, \cdots, A_k$ 构成 Ω 的分割 (即 $A_1, \cdots, A_k$ 两两不相容, 且其和是 Ω; 见 §1 第 3 小节).

记 $p_i = \mathbf{P}\{\xi = x_i\}$. 直观上显然, 如果在 n 次独立重复试验中观测随机变量 ξ 的取值, 则取 x_i 的值大致应该出现 $np_i (i = 1, \cdots, k)$ 次. (最好将这一段话与大数定

律的论点对比, 关于大数定律见 §5 和 §12.) 因此, 根据 n 次试验的结果, 计算的该随机变量的 "平均值" 大致为

$$\frac{1}{n}[np_1x_1+\cdots+np_kx_k]=\sum_{i=1}^{k}p_ix_i.$$

这一事实引出下面的定义.

定义 4 实数

$$\mathbf{E}\xi=\sum_{i=1}^{k}x_i\mathbf{P}(A_i) \tag{6}$$

称做随机变量

$$\xi=\sum_{i=1}^{k}x_iI(A_i)$$

的数学期望或平均值. 由于 $A_i=\{\omega:\xi(\omega)=x_i\}$, 而 $P_\xi(x_i)=\mathbf{P}(A_i)$, 则

$$\mathbf{E}\xi=\sum_{i=1}^{k}x_iP_\xi(x_i). \tag{7}$$

注意到分布函数的定义 $F_\xi=F_\xi(x)$, 并记

$$\Delta F_\xi(x)=F_\xi(x)-F_\xi(x-),$$

得 $P_\xi(x_i)=\Delta F_\xi(x_i)$, 从而

$$\mathbf{E}\xi=\sum_{i=1}^{k}x_i\Delta F_\xi(x_i). \tag{8}$$

在研究数学期望的性质之前, 常需要随机变量 ξ 的各种不同表达形式, 例如

$$\xi(\omega)=\sum_{j=1}^{l}x'_jI(B_j),$$

其中 $B_1+\cdots+B_l=\Omega$, 但是一般在 x'_j 之中可能有相同的值. 这时, 可以按上面的公式计算数学期望, 而不需要首先变换成所有的 x_i 值两两不等的 (5) 式. 事实上

$$\sum_{\{j:x'_j=x_i\}}x'_j\mathbf{P}(B_j)=x_i\sum_{\{j:x'_j=x_i\}}\mathbf{P}(B_j)=x_i\mathbf{P}(A_i),$$

于是

$$\sum_{j=1}^{l}x_jP(B_j)=\sum_{i=1}^{k}x_iP(A_i).$$

5. 数学期望的基本性质 现在列举数学期望的基本性质:

1) 若 $\xi \geqslant 0$, 则 $\mathbf{E}\xi \geqslant 0$.

2) $\mathbf{E}(a\xi + b\eta) = a\mathbf{E}\xi + b\mathbf{E}\eta$, 其中 a, b 是常数.

3) 若 $\xi \geqslant \eta$, 则 $\mathbf{E}\xi \geqslant \mathbf{E}\eta$.

4) $|\mathbf{E}\xi| \leqslant \mathbf{E}|\xi|$.

5) 若 ξ 和 η 独立, 则 $\mathbf{E}\xi\eta = \mathbf{E}\xi \cdot \mathbf{E}\eta$.

6) $(\mathbf{E}|\xi\eta|)^2 \leqslant \mathbf{E}\xi^2 \cdot \mathbf{E}\eta^2$ (柯西 – 布尼亚科夫斯基不等式, 亦称柯西 – 施瓦茨不等式, 或施瓦茨不等式).

7) 若 $\xi = I(A)$, 则 $\mathbf{E}\xi = \mathbf{P}(A)$.

性质 1) 和 7) 显然. 为证明性质 2), 设

$$\xi = \sum x_i I(A_i), \quad \eta = \sum y_j I(B_j),$$

则

$$\begin{aligned} a\xi + b\eta &= a\sum_{i,j} x_i I(A_i \cap B_j) + b\sum_{i,j} y_j I(A_i \cap B_j) \\ &= \sum_{i,j} (ax_i + by_j) I(A_i \cap B_j); \\ \mathbf{E}(a\xi + b\eta) &= \sum_{i,j} (ax_i + by_j) \mathbf{P}(A_i \cap B_j) \\ &= \sum_i ax_i \mathbf{P}(A_i) + \sum_j by_j \mathbf{P}(B_j) \\ &= a\sum_i x_i \mathbf{P}(A_i) + b\sum_j y_j \mathbf{P}(B_j) = a\mathbf{E}\xi + b\mathbf{E}\eta. \end{aligned}$$

由性质 1) 和 2), 可以证明 3). 因为

$$|\mathbf{E}\xi| = \left|\sum_i x_i \mathbf{P}(A_i)\right| \leqslant \sum_i |x_i| \mathbf{P}(A_i) = \mathbf{E}|\xi|,$$

所以性质 4) 显然. 为证明性质 5), 只需注意到

$$\begin{aligned} \mathbf{E}\xi\eta &= \mathbf{E}\left(\sum_i x_i I(A_i)\right)\left(\sum_j y_j I(B_j)\right) = \mathbf{E}\sum_{i,j} x_i y_j I(A_i \cap B_j) \\ &= \sum_{i,j} x_i y_j \mathbf{P}(A_i \cap B_j) = \sum_{i,j} x_i y_j \mathbf{P}(A_i)\mathbf{P}(B_j) \\ &= \left(\sum_i x_i \mathbf{P}(A_i)\right)\left(\sum_j y_j \mathbf{P}(B_j)\right) = \mathbf{E}\xi \cdot \mathbf{E}\eta, \end{aligned}$$

其中在证明过程中用到: 对于独立随机变量 ξ 和 η, 事件

$$A_i = \{\omega : \xi(\omega) = x_i\} \text{ 和 } B_j = \{\omega : \eta(\omega) = y_j\}$$

独立: $\mathbf{P}(A_i \bigcap B_j) = \mathbf{P}(A_i)\mathbf{P}(B_j)$.

为证明性质 6), 注意到

$$\xi^2 = \sum_{i=1}^{l} x_i^2 I(A_i), \quad \eta^2 = \sum_{j=1}^{k} y_j^2 I(B_j)$$

和

$$\mathbf{E}\xi^2 = \sum_{i=1}^{l} x_i^2 \mathbf{P}(A_i), \quad \mathbf{E}\eta^2 = \sum_{j=1}^{k} y_j^2 \mathbf{P}(B_j).$$

设 $\mathbf{E}\xi^2 > 0, \mathbf{E}\eta^2 > 0$. 记

$$\widetilde{\xi} = \frac{\xi}{\sqrt{\mathbf{E}\xi^2}}, \quad \widetilde{\eta} = \frac{\eta}{\sqrt{\mathbf{E}\eta^2}}.$$

由 $2|\widetilde{\xi}\widetilde{\eta}| \leqslant \widetilde{\xi}^2 + \widetilde{\eta}^2$, 可见 $2\mathbf{E}|\widetilde{\xi}\widetilde{\eta}| \leqslant \mathbf{E}\widetilde{\xi}^2 + \mathbf{E}\widetilde{\eta}^2 = 2$. 因此

$$\mathbf{E}|\widetilde{\xi}\widetilde{\eta}| \leqslant 1, \quad (\mathbf{E}|\xi\eta|)^2 \leqslant \mathbf{E}\xi^2 \times \mathbf{E}\eta^2.$$

假如 $\mathbf{E}\xi^2 = 0$, 则

$$\sum_i x_i^2 \mathbf{P}\{A_i\} = 0,$$

从而 0 是 ξ 的可能值, 并且

$$\mathbf{P}\{\omega : \xi(\omega) = 0\} = 1.$$

因此, 如果 $\mathbf{E}\xi^2$ 或 $\mathbf{E}\eta^2$ 之一等于 0, 则显然 $\mathbf{E}|\xi\eta| = 0$. 于是柯西 – 布尼亚科夫斯基不等式仍然成立.

注 性质 5) 明显地可以推广到任意有穷个随机变量: 若 $\xi_1, \cdots, \xi_r$独立, 则

$$\mathbf{E}\xi_1 \cdots \xi_r = \mathbf{E}\xi_1 \cdots \mathbf{E}\xi_r.$$

这里可以仿照 $n = 2$ 的情形证明, 亦可用归纳法证明.

例 3 设 ξ 是伯努利随机变量, 以概率 p 和 q 取 1 和 0 为值, 则

$$\mathbf{E}\xi = 1 \times \mathbf{P}\{\xi = 1\} + 0 \times \mathbf{P}\{\xi = 0\} = p.$$

例 4 设 $\xi_1, \cdots, \xi_n$ 是 n 个伯努利随机变量, 以概率 $\mathbf{P}\{\xi_i = 1\} = p$ 和 $\mathbf{P}\{\xi_i = 0\} = q, p + q = 1$ 取 1 和 0 为值, 则对于

$$S_n = \xi_1 + \cdots + \xi_n,$$

其数学期望为

$$\mathbf{E}S_n = np.$$

这里有另一种求法. 易见, 如果假设 $\xi_1, \cdots, \xi_n$ 独立伯努利随机变量, 则 $\mathbf{E}S_n$ 不变. 在此条件下, 根据 (4) 式, 有

$$\mathbf{P}\{S_n = k\} = \mathrm{C}_n^k p^k q^{n-k} \quad (k = 0, 1, \cdots, n).$$

因此,

$$\begin{aligned}\mathbf{E}S_n &= \sum_{k=0}^{n} k\mathbf{P}\{S_n = k\} = \sum_{k=0}^{n} k\mathrm{C}_n^k p^k q^{n-k} = \sum_{k=0}^{n} \frac{kn!}{k!(n-k)!} p^k q^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} p^{k-1} q^{(n-1)-(k-1)} \\ &= np \sum_{l=0}^{n-1} \frac{(n-1)!}{l![(n-1)-l]!} p^l q^{(n-1)-l} = np.\end{aligned}$$

其实, 用前一种方法比用后一种方法得到结果更快一些.

6. 随机变量函数的数学期望 设 $\xi = \sum x_i I(A_i)$, 其中 $A_i = \{\omega : \xi(\omega) = x_i\}$, 而 $\varphi = \varphi(\xi(\omega))$ 是 $\xi(\omega)$ 的某一函数. 如果 $B_j = \{\omega : \varphi(\xi(\omega)) = y_j\}$, 则 (设 $I_{B_j}(\omega) = I(B_j)$)

$$\varphi(\xi(\omega)) = \sum_j y_j I_{B_j}(\omega),$$

从而

$$\mathbf{E}\varphi = \sum_j y_j \mathbf{P}(B_j) = \sum_j y_j P_\varphi(y_j). \tag{9}$$

同样显然 (设 $I_{A_i}(\omega) = I(A_i)$)

$$\varphi(\xi(\omega)) = \sum_i \varphi(x_i) I_{A_i}(\omega).$$

于是, 为求 $\varphi = \varphi(\xi(\omega))$ 的数学期望, 既可以利用 (9) 式, 也可以利用下面的公式

$$\mathbf{E}\varphi(\xi(\omega)) = \sum_i \varphi(x_i) P_\xi(x_i).$$

7. 方差和标准差 随机变量 ξ 的方差和标准差, 表征 ξ 取值的散布程度, 是十分重要的概念.

定义 5 称

$$\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2$$

为随机变量的方差 (记作 $\mathbf{D}\xi$). 称 $\sigma = \sqrt{\mathbf{D}\xi}$ 为标准差.

由于

$$\mathbf{E}(\xi-\mathbf{E}\xi)^2=\mathbf{E}[\xi-2\xi\mathbf{E}\xi+(\mathbf{E}\xi)^2]=\mathbf{E}\xi^2-(\mathbf{E}\xi)^2,$$

可见

$$\mathbf{D}\xi=\mathbf{E}\xi^2-(\mathbf{E}\xi)^2.$$

显然 $\mathbf{D}\xi\geqslant 0$. 由方差的定义, 可见对于任意常数 a,b,

$$\mathbf{D}(a+b\xi)=b^2\mathbf{D}\xi,$$

特别 $\mathbf{D}a=0,\mathbf{D}(b\xi)=b^2\mathbf{D}\xi$.

对于二随机变量 ξ 和 η,

$$\begin{aligned}\mathbf{D}(\xi+\eta)&=\mathbf{E}[(\xi-\mathbf{E}\xi)+(\eta-\mathbf{E}\eta)]^2\\&=\mathbf{D}\xi+\mathbf{D}\eta+2\mathbf{E}(\xi-\mathbf{E}\xi)(\eta-\mathbf{E}\eta).\end{aligned}$$

记

$$\operatorname{cov}(\xi,\eta)=\mathbf{E}(\xi-\mathbf{E}\xi)(\eta-\mathbf{E}\eta),$$

称做随机变量 ξ 和 η 的*协方差*. 如果 $\mathbf{D}\xi\geqslant 0,\mathbf{D}\eta\geqslant 0$, 则

$$\rho(\xi,\eta)=\frac{\operatorname{cov}(\xi,\eta)}{\sqrt{\mathbf{D}\xi\times\mathbf{D}\eta}}$$

称做随机变量 ξ 和 η 的*相关系数*. 不难证明 (见练习题 7), 若 $\rho(\xi,\eta)=\pm1$, 则随机变量 ξ 和 η 线性相关:

$$\eta=a\xi+b,$$

其中当 $\rho(\xi,\eta)=1$ 时 $a>0$; 当 $\rho(\xi,\eta)=-1$ 时 $a<0$.

立即可以指出, 若 ξ 和 η 独立, 则 $\xi-\mathbf{E}\xi$ 和 $\eta-\mathbf{E}\eta$ 独立, 因此根据数学期望的性质 5), 有

$$\operatorname{cov}(\xi,\eta)=\mathbf{E}(\xi-\mathbf{E}\xi)\times\mathbf{E}(\eta-\mathbf{E}\eta)=0.$$

由协方差的定义, 可见

$$\mathbf{D}(\xi+\eta)=\mathbf{D}\xi+\mathbf{D}\eta+2\operatorname{cov}(\xi,\eta).\tag{10}$$

如果 *ξ 与 η 独立*, 则*和 $\xi+\eta$ 的方差等于方差之和*:

$$\mathbf{D}(\xi+\eta)=\mathbf{D}\xi+\mathbf{D}\eta.\tag{11}$$

虽然式 (11) 可以由式 (10) 导出的, 然而 (11) 式在比 "ξ 和 η 独立" 较弱的条件下仍然成立. 具体地说, 只要假设 "*ξ 和 η 不相关*", 即假设 $\operatorname{cov}(\xi,\eta)=0$ 就可以了.

注 由 ξ 和 η 不相关, 一般得不出 ξ 和 η 独立. 看下面的例子. 假设随机变量 α 以概率 1/3 分别取 $0, \pi/2, \pi$ 为值, 则 $\xi = \sin\alpha$ 和 $\eta = \cos\alpha$ 不相关, 然而 ξ 和 η 不但 (关于 $\mathbf{P}$) 不独立:

$$\mathbf{P}\{\xi = 1, \eta = 1\} = 0 \neq 1/9 = \mathbf{P}\{\xi = 1\}\mathbf{P}\{\eta = 1\},$$

而且 ξ 和 η 之间有*函数关系*: $\xi^2 + \eta^2 = 1$.

性质 (10) 和 (11) 明显可以推广到任意个随机变量 $\xi_1, \cdots, \xi_n$ 的情形:

$$\mathbf{D}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \mathbf{D}\xi_i + 2\sum_{i>j} \operatorname{cov}(\xi_i, \xi_j). \tag{12}$$

特别, 若随机变量 $\xi_1, \cdots, \xi_n$ *两两独立* (实际上, 只需要求它们*两两不相关*), 则

$$\mathbf{D}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \mathbf{D}\xi_i. \tag{13}$$

例 5 设 ξ 是伯努利随机变量, 以概率 p 和 q 取 1 和 0 为值, 则

$$\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2 = \mathbf{E}(\xi - p)^2 = (1-p)^2 p + p^2 q = pq.$$

由此可见, 如果 $\xi_1, \cdots, \xi_n$ 是独立同分布的伯努利随机变量序列, 且 $S_n = \xi_1 + \cdots + \xi_n$, 则

$$\mathbf{D}S_n = npq. \tag{14}$$

8. 最优线性估计 考虑两个随机变量 ξ 和 η. 假设只对随机变量 ξ 进行观测. 如果随机变量 ξ 和 η 相关, 则可以预期, 已知 ξ 的值可以对未观测随机变量 η 的值, 作出某种判断.

我们把 ξ 的任何一个函数 $f = f(\xi)$ 称做 η 的一个*估计量*. 称估计量 $f^* = f^*(\xi)$ 为*在均方意义下最优的*, 如果

$$\mathbf{E}(\eta - f^*(\xi))^2 = \inf_f \mathbf{E}(\eta - f(\xi))^2.$$

现在讨论, 如何在*线性估计* $\lambda(\xi) = a + b\xi$ 类中求*最优估计*. 为此考虑函数

$$g(a,b) = \mathbf{E}[\eta - (a + b\xi)]^2.$$

将 $g(a,b)$ 分别对 a 和 b 求偏导数, 得

$$\frac{\partial g(a,b)}{\partial a} = -2\mathbf{E}[\eta - (a + b\xi)],$$
$$\frac{\partial g(a,b)}{\partial a} = -2\mathbf{E}[\eta - (a + b\xi)\xi].$$

令所得偏导数等于 0, 可以求出 λ 均方线性估计: $\lambda^* = a^* + b^*\xi$, 其中

$$a^* = \mathbf{E}\eta - b^*\mathbf{E}\xi, \quad b^* = \frac{\operatorname{cov}(\xi,\eta)}{\mathbf{D}\xi}, \tag{15}$$

即

$$\lambda^*(\xi) = \mathbf{E}\eta + \frac{\operatorname{cov}(\xi,\eta)}{\mathbf{D}\xi}(\xi - \mathbf{E}\xi). \tag{16}$$

称 $\mathbf{E}[\eta - \lambda^*(\xi)]^2$ 为估计值的*均方误差*. 经过简单的计算可见, 该误差等于

$$\Delta^* = \mathbf{E}[\eta - \lambda^*(\xi)]^2 = \mathbf{D}\eta - \frac{\operatorname{cov}^2(\xi,\eta)}{\mathbf{D}\xi} = \mathbf{D}\eta \times [1 - \rho^2(\xi,\eta)]. \tag{17}$$

这样, ξ 和 η 之间的*相关系数* (绝对值)越大, *均方误差*的估计值 Δ^* 就越小. 特别, 如果 $|\rho(\xi,\eta)| = 1$, 则 $\Delta^* = 0$ (与练习题 7 的结果比较). 如果随机变量 ξ 和 η 不相关 (即 $\rho(\xi,\eta) = 0$), 则 $\lambda^*(\xi) = \mathrm{E}\eta$. 于是, 在随机变量 ξ 和 η 不相关的情形下, 根据 ξ 对 η 的估计就是 $\mathrm{E}\eta$ (与练习题 4 的结果比较).

9. 练习题

1. 验证示性函数 $I_A = I_A(\omega)$ 的下列性质:

$$\begin{aligned}
&I_\varnothing = 0, \quad I_\Omega = 1, \quad I_{\overline{A}} = 1 - I_A, \\
&I_{AB} = I_A \times I_B, \quad I_{A\cup B} = I_A + I_B - I_{AB}, \\
&I_{A\setminus B} = I_A(1 - I_B), \quad I_{A\triangle B} = (I_A - I_B)^2 = I_A + I_B (\mathrm{mod} 2), \\
&I_{E_1} = 1 - \prod_{i=1}^n (1 - I_{A_i}), \quad I_{E_2} = \prod_{i=1}^n (1 - I_{A_i}), \quad I_{E_3} = \sum_{i=1}^n I_{A_i},
\end{aligned}$$

其中 $A\triangle B$ 称做集合 A 与 B 的*对称差*: $A\triangle B = (A\setminus B)\bigcup(B\setminus A)$, 而

$$E_1 = \bigcup_{i=1}^n A_i, \quad E_2 = \overline{\bigcup_{i=1}^n A_i}, \quad E_3 = \sum_{i=1}^n A_i.$$

2. 设是 $\xi_1, \cdots, \xi_n$ 独立随机变量, 且

$$\xi_{\min} = \min\{\xi_1, \cdots, \xi_n\}, \quad \xi_{\max} = \max\{\xi_1, \cdots, \xi_n\}.$$

证明:

$$\mathbf{P}\{\xi_{\min} \geqslant x\} = \prod_{i=1}^n \mathbf{P}\{\xi_i \geqslant x\}, \quad \mathbf{P}\{\xi_{\max} < x\} = \prod_{i=1}^n \mathbf{P}\{\xi_i < x\}.$$

3. 设 $\xi_1, \cdots, \xi_n$ 是独立伯努利随机变量, 且

$$\begin{aligned}
\mathbf{P}\{\xi_i = 0\} &= 1 - \lambda_i\Delta, \\
\mathbf{P}\{\xi_i = 1\} &= \lambda_i\Delta,
\end{aligned}$$

其中 $\Delta > 0, \lambda_i > 0$, 而 Δ 是较小的数. 证明:

$$\mathbf{P}\{\xi_1 + \cdots + \xi_n = 1\} = \left(\sum_{i=1}^{n} \lambda_i\right)\Delta + O(\Delta^2),$$
$$\mathbf{P}\{\xi_1 + \cdots + \xi_n > 1\} = O(\Delta^2).$$

4. 证明, 当 $a = \mathbf{E}\xi$ 时 $\mathbf{E}(\xi - a)^2$ 达到下确界

$$\inf_{-\infty<a<\infty} \mathbf{E}(\xi - a)^2, \quad 即 \quad \inf_{-\infty<a<\infty} \mathbf{E}(\xi - a)^2 = \mathbf{D}\xi.$$

5. 设 $F_\xi(x)$ 是随机变量 ξ 的分布函数, 而 m_e 是 $F_\xi(x)$ 的中位数, 即下列条件的点:

$$F_\xi(m_e-) \leqslant \frac{1}{2} \leqslant F_\xi(m_e).$$

证明

$$\inf_{-\infty<a<\infty} \mathbf{E}|\xi - a| = \mathbf{E}|\xi - m_e|.$$

6. 设 $P_\xi(x) = \mathbf{P}\{\xi = x\}, F_\xi(x) = \mathbf{P}\{\xi \leqslant x\}$, 证明:

(1) 对于 $a > 0, -\infty < b < \infty$, 有

$$P_{a\xi+b}(x) = P_\xi\left(\frac{x-b}{a}\right),$$
$$F_{a\xi+b}(x) = F_\xi\left(\frac{x-b}{a}\right).$$

(2) 如果 $y \geqslant 0$, 则

$$F_{\xi^2}(y) = F_\xi(+\sqrt{y}) - F_\xi(-\sqrt{y}) + P_\xi(-\sqrt{y}).$$

(3) 设 $\xi^+ = \max\{\xi, 0\}$, 则

$$F_{\xi^+}(x) = \begin{cases} 0, & 若\ x < 0, \\ F_\xi(x), & 若\ x = 0, \\ F_\xi(x), & 若\ x > 0. \end{cases}$$

7. 设随机变量 ξ 和 η 的方差 $\mathbf{D}\xi > 0, \mathbf{D}\eta > 0$, 它们的相关系数为 $\rho = \rho(\xi, \eta)$. 证明:

(1) $|\rho| \leqslant 1$.

(2) 若 $|\rho| = 1$, 则存在常数 a 和 b, 使 $\eta = a\xi + b$, 并且当 $\rho = 1$ 时

$$\frac{\eta - \mathbf{E}\eta}{\sqrt{\mathbf{D}\eta}} = \frac{\xi - \mathbf{E}\xi}{\sqrt{\mathbf{D}\xi}},$$

(即 $a > 0$), 当 $\rho = -1$ 时

$$\frac{\eta - \mathbf{E}\eta}{\sqrt{\mathbf{D}\eta}} = -\frac{\xi - \mathbf{E}\xi}{\sqrt{\mathbf{D}\xi}},$$

(即 $a<0$).

8. 假设对于随机变量 ξ 和 $\eta, \mathbf{E}\xi=\mathbf{E}\eta=0, \mathbf{D}\xi=\mathbf{D}\eta=1$, 而 ξ 和 η 的相关系数为 $\rho=\rho(\xi,\eta)$. 证明:

$$\mathbf{E}\max\{\xi^2,\eta^2\}\leqslant 1+\sqrt{1-\rho^2}.$$

9. 利用等式

$$I_{E_2}=\prod_{i=1}^{n}(1-I_{A_i}),\quad 其中\quad E_2=\overline{\bigcup_{i=1}^{n}A_i},$$

证明 §1 中练习题 5 的公式:

$$\mathbf{P}(B_0)=1-S_1+S_2-\cdots\pm S_n.$$

10. 设 $\xi_1,\cdots,\xi_n$ 是独立随机变量, $\varphi_1=\varphi_1(\xi_1,\cdots,\xi_k)$ 和 $\varphi_2=\varphi_2(\xi_{k+1},\cdots,\xi_n)$ 分别是 $(\xi_1,\cdots,\xi_k)$ 和 $(\xi_{k+1},\cdots,\xi_n)$ 的函数, 证明 φ_1 和 φ_2 独立.

11. 证明随机变量 $\xi_1,\cdots,\xi_n$ 独立, 当且仅当对于一切 $x_1,\cdots,x_n$,

$$F_{\xi_1,\cdots,\xi_n}(x_1,\cdots,x_n)=F_{\xi_1}(x_1)\cdots F_{\xi_n}(x_n),$$

其中 $F_{\xi_1,\cdots,\xi_n}(x_1,\cdots,x_n)=\mathbf{P}\{\xi_1\leqslant x_1,\cdots,\xi_n\leqslant x_n\}$.

12. 证明随机变量 ξ 与自己独立 (即 ξ 与 ξ 独立), 当且仅当 $\xi\equiv\text{const.}$ (常数).

13. 问随机变量 ξ 满足何条件时, ξ 与 $\sin\xi$ 独立?

14. 设 ξ 和 η 是独立随机变量, 且 $\eta\neq 0$. 通过概率 $P_\xi(x)$ 和 $P_\eta(y)$ 表示概率:

$$\mathbf{P}\{\xi\eta\leqslant z\}\quad 和\quad \mathbf{P}\left\{\frac{\xi}{\eta}\leqslant z\right\}.$$

15. 设随机变量 ξ,η,ζ 满足条件: $|\xi|\leqslant 1, |\eta|\leqslant 1, |\zeta|\leqslant 1$, 证明贝尔 (A. G. Bell) 不等式:

$$|\mathbf{E}\xi\zeta-\mathbf{E}\eta\zeta|\leqslant 1-\mathbf{E}\xi\eta.$$

(例如, 见 [136])

16. 向 n 个箱子中独立地掷 k 个球. 假设每个球落入各箱的概率都等于 $1/n$, 求非空箱子个数的数学期望.

§5. 伯努利概型 I. 大数定律

1. 伯努利概型 以上的定义 "三对象"

$$(\Omega,\mathscr{A},\mathbf{P}),\ 其中\ \Omega=\{\omega:\omega=(a_1,\cdots,a_n), a_i=0,1\},$$
$$\mathscr{A}=\{A:A\subseteq\Omega\},\quad \mathbf{P}(\{\omega\})=p^{\sum a_i}(1-p)^{n-\sum a_i}(=p(\omega)),$$

称做伯努利概型, 全称为 "有两种结局的 n 次独立试验的概率模型". 在这一节和下一节, 我们将研究 (在下面所指的意义下) 伯努利概型的某些性质. 用与之相联系的、随机变量和事件概率的术语, 引进这些性质较为适宜.

引进随机变量 $\xi_1, \cdots, \xi_n$, 其中 $\xi_i = \xi_i(\omega) = a_i, i = 1, \cdots, n$, 而 $\omega = (a_1, \cdots, a_n)$. 已经熟知, 伯努利随机变量 $\xi_i = \xi_i(\omega) = a_i, i = 1, \cdots, n$ 独立而且同分布:

$$\mathbf{P}\{\xi_i = 1\} = p, \mathbf{P}\{\xi_i = 0\} = 1 - p = q, \quad i = 1, \cdots, n.$$

随机变量 ξ_i 表示在第 i 步 (或时刻 i) 的试验结果.

设 $S_0(\omega) \equiv 0$,

$$S_k = \xi_1 + \cdots + \xi_k, \quad k = 1, \cdots, n.$$

我们已经 (在 §4 例 4) 证明, $\mathbf{E}S_n = np$, 从而

$$\mathbf{E}\frac{S_n}{n} = p. \tag{1}$$

即 "成功" 频率 S_n/n 的平均值等于成功的概率 p. 由此自然产生一个问题: "成功" 频率 S_n/n 对成功概率 p 的 (绝对) 偏差的大小如何?

我们首先指出, 对于充分小的 $\varepsilon > 0$ 和甚至很大的 n, 也不能指望对于任意 ω, 频率 S_n/n 对成功概率 p 的 (绝对) 偏差都小于 ε, 即不能指望对于任意 ω, 不等式

$$\left|\frac{S_n(\omega)}{n} - p\right| \leqslant \varepsilon, \quad \omega \in \Omega \tag{2}$$

都成立.

事实上, 对于 $0 < p < 1$, 由

$$\mathbf{P}\left\{\frac{S_n}{n} = 1\right\} = \{\xi_1 = 1, \cdots, \xi_n = 1\} = p^n,$$
$$\mathbf{P}\left\{\frac{S_n}{n} = 0\right\} = \{\xi_1 = 0, \cdots, \xi_n = 0\} = q^n,$$

可见, 对于充分小的 $\varepsilon > 0$ 不等式 (2) 并不成立.

不过, 我们指出当 n 很大时, 事件

$$\left\{\frac{S_n}{n} = 1\right\} \quad 和 \quad \left\{\frac{S_n}{n} = 0\right\}$$

的概率都较小. 因此, 自然想到, 当 n 充分大时使

$$\left|\frac{S_n}{n} - p\right| > \varepsilon$$

成立的 "结局 ω 的全体" 的概率也较小. 因此, 设法估计事件

$$\left\{\omega : \left|\frac{S_n(\omega)}{n} - p\right| > \varepsilon\right\}$$

的概率. 为此我们运用如下切比雪夫不等式.①

切比雪夫 (П. Л. Чебышёв) **不等式** 设 $(\Omega, \mathscr{A}, \mathbf{P})$ 是某一概率空间, $\xi = \xi(\omega)$ 是非负随机变量. 那么, 对任意 $\varepsilon > 0$,

$$\mathbf{P}\{\xi \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}\xi}{\varepsilon}. \tag{3}$$

证明 注意到.

$$\xi = \xi I(\xi \geqslant \varepsilon) + \xi I(\xi < \varepsilon) \geqslant \xi I(\xi \geqslant \varepsilon) \geqslant \varepsilon I(\xi \geqslant \varepsilon),$$

其中 $I(A)$ 是集合 A 的示性函数.

于是, 根据数学期望的性质

$$\mathbf{E}\xi \geqslant \varepsilon \mathbf{E} I(\xi \geqslant \varepsilon) = \varepsilon \mathbf{P}\{\xi \geqslant \varepsilon\},$$

从而 (3) 式得证. □

系 设 ξ 是任意随机变量, 则对任意 $\varepsilon > 0$,

$$\begin{aligned} &\mathbf{P}\{|\xi| \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}|\xi|}{\varepsilon}, \\ &\mathbf{P}\{|\xi| \geqslant \varepsilon\} = \mathbf{P}\{\xi^2 \geqslant \varepsilon^2\} \leqslant \frac{\mathbf{E}\xi^2}{\varepsilon^2}, \\ &\mathbf{P}\{|\xi - \mathbf{E}\xi| \geqslant \varepsilon\} \leqslant \frac{\mathbf{D}\xi}{\varepsilon^2}. \end{aligned} \tag{4}$$

利用最后一个不等式, 设 $\xi = S_n/n$, 则由 §4 (14) 式, 有

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} \leqslant \frac{\mathbf{D}\left(\dfrac{S_n}{n}\right)}{\varepsilon^2} = \frac{\mathbf{D}S_n}{n^2\varepsilon^2} = \frac{npq}{n^2\varepsilon^2} = \frac{pq}{n\varepsilon^2}.$$

于是

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} \leqslant \frac{pq}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}. \tag{5}$$

由此可见, 当 n 充分大时, "成功" 频率 S_n/n 对 "成功" 概率 p 的 (绝对) 偏差大于 ε 的概率充分小.

对于一切 n 和 $1 \leqslant k \leqslant n$, 记

$$P_n(k) = \mathrm{C}_n^k p^k q^{n-k},$$

则

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} = \sum_{\{k:\left|\frac{k}{n}-p\right| \geqslant \varepsilon\}} P_n(k),$$

① П. Л. Чебышёв 是俄罗斯数学家, 按俄语语音应译为 "切贝绍夫" (见《俄语姓名译名手册》. 商务印书馆, 1982:《新俄汉数学词汇》, 科学出版社, 1988). 英语文献一般译为 P. L. Chebyshev. 我国有的文献按英语译音, 译为 "切比雪夫" 或 "契比雪夫". —— 译者

实际上证明了 (5) 式:

$$\sum_{\{k:\left|\frac{k}{n}-p\right|\geqslant\varepsilon\}} P_n(k) \leqslant \frac{pq}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}, \tag{6}$$

即我们用概率的方法证明了不等式 (5). 注意, 假如不用概率的方法, 而用分析的方法, 也可以证明此不等式.

由 (6) 可见

$$\sum_{\{k:\left|\frac{k}{n}-p\right|\geqslant\varepsilon\}} P_n(k) \to 0 (n\to\infty). \tag{7}$$

该命题可以用图形作如下解释. 图 6 是二项分布 $\{P_n(k), 0 \leqslant k \leqslant n\} (p = 1/2)$ 的示意图.

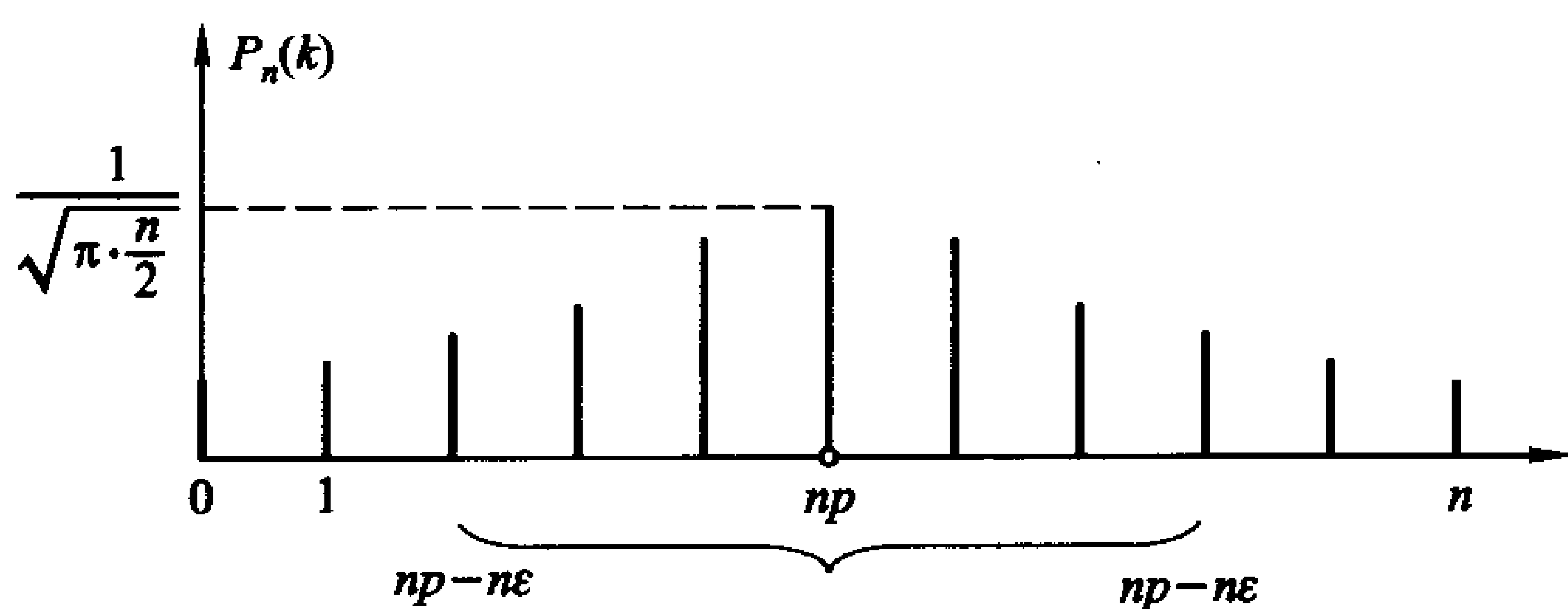


图 6 $(m_1 = np - n\varepsilon, m_2 = np + n\varepsilon)$

由图 6 可见: 概率 $P_n(k)$ 在 $k = np$ 处达到最大值 P_m, 其中

$$P_m = \frac{1}{\sqrt{\pi\cdot\frac{n}{2}}}.$$

图 6 显示: 若将概率 $P_n(k)$ 对 k 求和, 则对于 $np - n\varepsilon \leqslant k \leqslant np + n\varepsilon$, 概率接近 1.

我们把一系列随机变量 $S_0, S_1, \cdots, S_n$ 视为某游动的质点的轨道. 那么, 对 (7) 式可以作如下解释.

过原点引 3 条直线: $k(p-\varepsilon), kp, k(p+\varepsilon)$. 那么, 质点轨道总的趋势是沿直线运动. 对于任意 $\varepsilon > 0$, 可以断定, 对于充分大的 n, 表示质点在时刻 n 位置的点 S_n 位于区间 $[n(p-\varepsilon), n(p+\varepsilon)]$ 上 (见图 7).

命题 (7) 可以表示为:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty. \tag{8}$$

不过需要指出, 这里存在一定的细节. 问题在于, 假如概率 $\mathbf{P}$ 在某空间 $(\Omega, \mathscr{A})$ 上, 空间 $(\Omega, \mathscr{A})$ 上定义了独立无穷伯努利随机变量序列 $\xi_1, \xi_2, \cdots$, 则上面的写法是

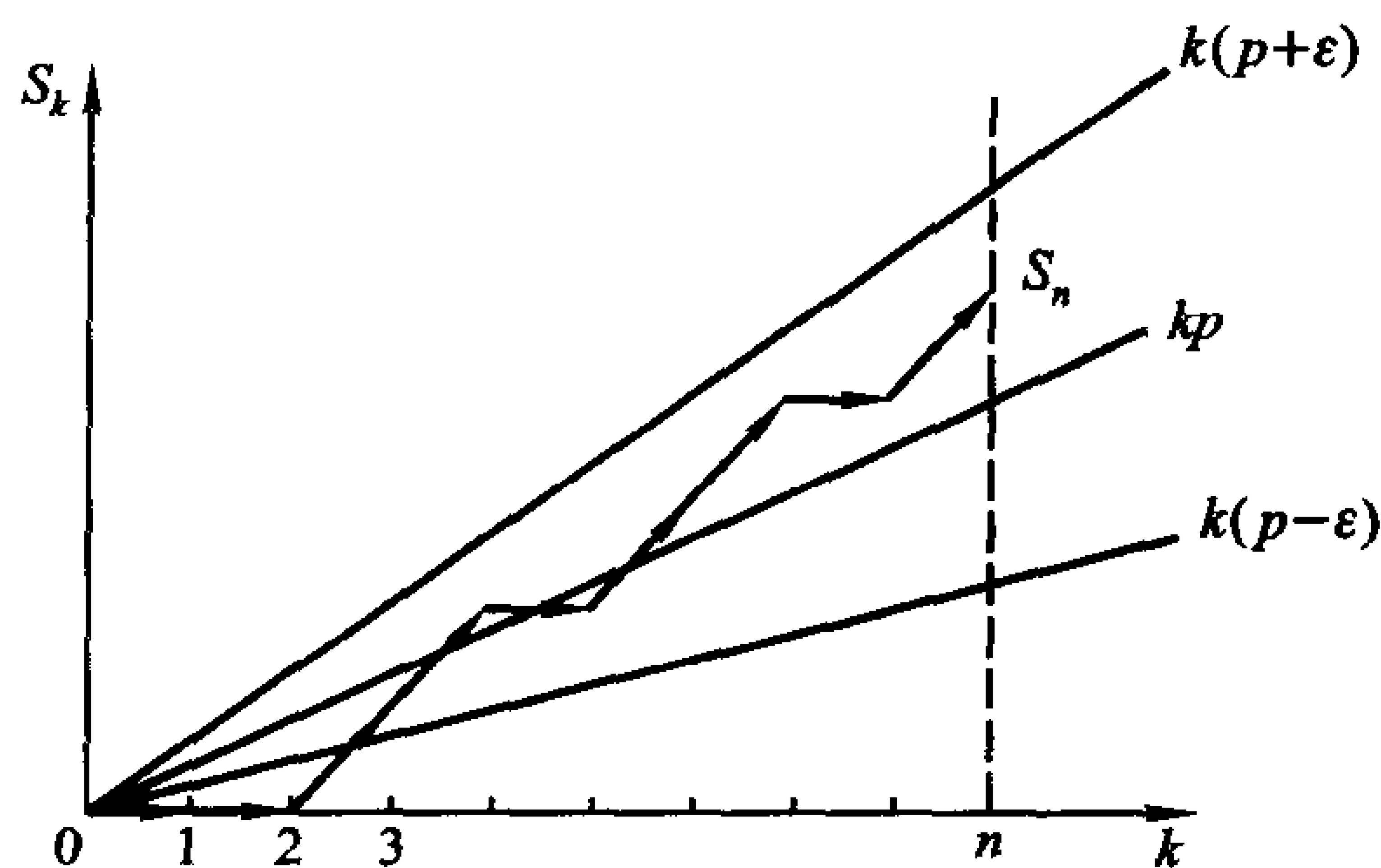


图 7

完全正确的: 确实可以建立这样的对象, 从而给 (8) 式赋予严格的概率意义 (见第二章 §9 定理 1 系 1). 现在, 如果希望赋予 (7) 式分析命题的含义, 则用初等概率的语言, 可以说明下述事实.

设 $(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathbf{P}^{(n)}), n \geqslant 1$, 是伯努利概型序列:

$$\Omega^{(n)} = \{\omega^{(n)} : \omega^{(n)} = (a_1^{(n)}, \cdots, a_n^{(n)}), a_i^{(n)} = 0, 1\},$$
$$\mathscr{A}^{(n)} = \{A : A \subseteq \Omega^{(n)}\},$$
$$\mathbf{P}^{(n)}(\{\omega^{(n)}\}) = p^{\sum a_i^{(n)}} q^{n-\sum a_i^{(n)}},$$

而

$$S_k^{(n)}(\omega^{(n)}) = (\xi_1^{(n)}(\omega^{(n)}), \cdots, \xi_k^{(n)}(\omega^{(n)})),$$

其中, 对于 $n \geqslant 1, \xi_1^{(n)}, \cdots, \xi_n^{(n)}$ 是独立同分布伯努利随机变量序列. 那么

$$\mathbf{P}^{(n)}\left\{\omega^{(n)} : \left|\frac{S_n^{(n)}(\omega^{(n)})}{n} - p\right| \geqslant \varepsilon\right\} = \sum_{\{k:\left|\frac{k}{n}-p\right|\geqslant\varepsilon\}} P_n(k) \to 0, \quad (n \to \infty). \tag{9}$$

式 (7) ~ (9) 的命题称做

“伯努利大数定律”

应该指出, J. 伯努利证明的恰好是命题 (7). 其证明非常严格, 利用了二项分布 “尾部” 概率的估计, 即对于满足 $|k/n - p| \geqslant \varepsilon$ 的 k 估计概率 $P_n(k)$. 对于充分大的 n, 二项分布 “尾部” 概率

$$\sum_{\{k:\left|\frac{k}{n}-p\right|\geqslant\varepsilon\}} P_n(k)$$

的直接计算问题相当繁杂, 况且所得 “频率 S_n/n 对概率 p 绝对偏差小于 ε” 的概率估计式很难实际应用. 因此, 对于任意 p, 棣莫弗和拉普拉斯所创造的概率 $P_n(k)$ 的

渐近公式特别重要, 不仅重新证明了大数定律, 而且得到了更精确的所谓局部及积分极限定理. 该定理的实质在于, 对于充分大 n 的和至少满足 $k \sim np$ 的 k, 有

$$P_n(k) \sim \frac{1}{\sqrt{2\pi npq}} \mathrm{e}^{-\frac{(k-np)^2}{2npq}},$$

而

$$\sum_{\left\{k:\left|\frac{k}{n}-p\right| \leqslant \varepsilon\right\}} P_n(k) \sim \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x.$$

2. 大数定律的意义 下一节将给出上述结果的确切表述和证明. 现在, 我们讨论大数定律的现实意义, 及其经验解释.

假设进行大量, 例如 N 系列试验, 而每一系列试验包括 "n 次独立试验, 而每次试验以概率 p 出现某事件 C". 设 S_n^i/n 是事件 C 在第 i 系列试验中出现的频率, N_ε 是 "频率对概率的绝对偏差不大于 ε" 系列数, 即

$$N_\varepsilon \text{等于使} \left|\frac{S_n^i}{n} - p\right| \leqslant \varepsilon \text{的 } i \text{ 个数}.$$

那么, 由大数定律可见

$$\frac{N_\varepsilon}{N} \sim P_\varepsilon, \tag{10}$$

其中

$$P_\varepsilon = \mathbf{P}\left\{\left|\frac{S_n^1}{n} - p\right| \leqslant \varepsilon\right\}.$$

这里, 重要的是强调, 将 (10) 式精确化的尝试无疑将必须利用某一概率测度, 像估计频率 S_n/n 对概率 p 的偏差一样, 这种估计只有在引进概率测度 $\mathbf{P}$ 后才有可能.

3. 观测次数 考虑上面得到的估计

$$\mathbf{P}\left\{\left|\frac{S_n^i}{n} - p\right| \geqslant \varepsilon\right\} = \sum_{\left\{k:\left|\frac{k}{n}-p\right| \geqslant \varepsilon\right\}} P_n(k) \leqslant \frac{1}{4n\varepsilon^2}. \tag{11}$$

为回答下面数理统计的典型问题: 对任意 $0 < p < 1$, 保证不等式

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \alpha \tag{12}$$

成立的最小观测次数 n 如何? 其中 α 是给定的通常较小的数.

由 (11) 式可见, 满足 (12) 式的最小观测次数, 是满足

$$n \geqslant \frac{1}{4\varepsilon^2 \alpha} \tag{13}$$

的最小整数 n.

例如, 若取 $\alpha = 0.05, \varepsilon = 0.02$, 则观测次数为 12 500 就可以满足 (12) 式, 而且不依赖于参数 p.

我们在下面 (§6, 第 5 小节) 将看到, 观测次数可以大为减小, 因为切比雪夫不等式作为概率

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\}$$

的上侧估计太粗略.

4. 熵　记

$$C(n,\varepsilon) = \left\{\omega : \left|\frac{S_n(\omega)}{n} - p\right| \leqslant \varepsilon\right\}.$$

由大数定律可见, 对于任何 $\varepsilon > 0$, 当 n 充分大时, 集合 $C(n,\varepsilon)$ 的概率接近 1. 这时, $C(n,\varepsilon)$ 中的轨道 (实现) ω 自然称做**典型的** [或 $C(n,\varepsilon)$- **典型的**].

提出下面的问题: **典型实现的条数** $N(C(n,\varepsilon))$ 以及**每一条典型实现的权** $p(\omega)$ 如何?

为此, 首先注意到, 基本事件空间 Ω 中点的总数 $N(\Omega) = 2^n$, 而对于 $p = 0$ 或 1, 典型轨道只有一条: $(1,1,\cdots,1)$ 或 $(0,0,\cdots,0)$, 即 $C(n,\varepsilon) = 1$. 但是, 假如 $p = 1/2$, 则直观上显然, "几乎一切" 轨道 [只有 $(1,1,\cdots,1)$ 或 $(0,0,\cdots,0)$ 除外] 都是典型的, 因而轨道的条数接近 2^n.

结果表明, 对于 $0 < p < 1$, 所提出问题有完全确定的答案: 无论是典型轨道数, 还是权重 $p(\omega)$, 都决定于 p 的某一专门函数 —— "熵".

为更深入地揭示相应结果的内容, 我们考虑 §2 第 2 小节中比伯努利概型更加一般的概型.

设 $(p_1, p_2, \cdots, p_r)$ 是一有限概率分布, 即满足条件 $p_1 + p_2 + \cdots + p_r = 1$ 的非负实数. 称

$$H = -\sum_{i=1}^{r} p_i \ln\ p_i \tag{14}$$

为概率分布 $(p_1, p_2, \cdots, p_r)$ 的熵, 其中 ln 是自然对数, 且 $0\ln 0 = 0$. 显然, $H \geqslant 0$, 而且 $H = 0$ 当且仅当在 $p_1, p_2, \cdots, p_r$ 中除某一个为 1 之外都等于 0. 函数 $f(x) = -x\ln x (0 < x < 1)$ 是 (向上) 凸函数. 熟知, 由凸函数的性质, 有

$$\frac{f(x_1) + \cdots + f(x_r)}{r} \leqslant f\left(\frac{x_1 + \cdots + x_r}{r}\right).$$

从而

$$H = -\sum_{i=1}^{r} p_i \ln p_i \leqslant -r \times \frac{p_1 + \cdots + p_r}{r} \times \ln\left(\frac{p_1 + \cdots + p_r}{r}\right) = \ln r.$$

换句话说, 当 $p_1 = p_2 = \cdots = p_r = 1/r$ 时, 熵达到其最大值 (对于 $r = 2$, 函数 $H = H(p)$ 的图形见图 8).

如果把 $p_1, p_2, \cdots, p_r$ 看成某些事件, 如 $A_1, A_2, \ldots, A_r$ 出现的概率, 则完全清楚: 出现某个事件 "不确定性的程度", 对于不同的分布是不同的. 例如, $p_1 = 1, p_2 = \cdots = p_r = 0$, 则该分布不具有任何不确定性: 可以满怀信心地说, 试验结果必然出现事件 A_1. 不过, 如果 $p_1 = p_2 = \cdots = p_r = 1/r$, 则这样的分布具有最大的不确定性, 因为甚至不能说哪个事件出现的可能性更大些.

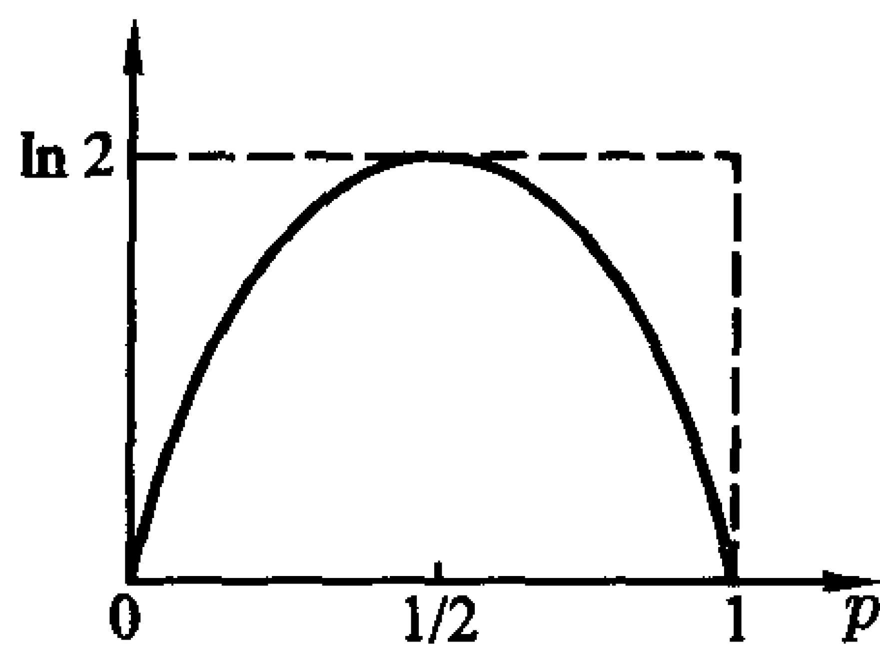


图 8 函数 $H(p)$

为比较不同分布的不确定性, 需要有不同分布的不确定性之**度量的数字特征**. 熵 H 正是不确定性度量的恰当的数字特征. 由下面的讨论可见, 熵在统计力学、编码理论和通讯理论中起着重要作用.

假设

$$\Omega = \{\omega : \omega = (a_1, \cdots, a_n), a_i = 1, \cdots, r\}$$

是基本事件空间, 其中 $p(\omega) = p_1^{\nu_1(\omega)} \cdots p_r^{\nu_r(\omega)}$, 其中 $\nu_i(\omega)$ 是序列 ω 中第 i 个元素 a_i 的个数, 而 $(p_1, \cdots, p_r)$ 是某一概率分布.

对于任意 $\varepsilon > 0$, 设

$$C(n, \varepsilon) = \left\{\omega : \left|\frac{\nu_i(\omega)}{n} - p_i\right| < \varepsilon, i = 1, \cdots, r\right\}.$$

显然,

$$\mathbf{P}(C(n, \varepsilon)) \geqslant 1 - \sum_{i=1}^{r} \mathbf{P}\left\{\left|\frac{\nu_i(\omega)}{n} - p_i\right| \geqslant \varepsilon\right\},$$

并且, 由于大数定律知, 该式也适用于随机变量

$$\xi_k(\omega) = \begin{cases} 1, & 若\ a_k = i, \\ 0, & 若\ a_k \neq i, \end{cases} \quad k = 1, \cdots, n,$$

概率

$$\mathbf{P}\left\{\left|\frac{\nu_i(\omega)}{n} - p_i\right| \geqslant \varepsilon\right\}$$

充分地小. 因此对于充分大的 n, $C(n, \varepsilon)$ 的概率接近 1. 同 $r = 2$ 的情形一样, 进入 $C(n, \varepsilon)$ 的轨道称为**典型的**.

如果所有 $p_i > 0 (i = 1, \cdots, r)$, 则对于任何 $\omega \in \Omega$, 权重

$$p(\omega) = \exp\left\{-n \sum_{k=1}^{r} \left[-\frac{\nu_k(\omega)}{n} \ln p_k\right]\right\}.$$

因此, 如果 ω 是**典型轨道**, 则由 (14) 式, 有

$$\left|\sum_{i=1}^{r} \left[-\frac{\nu_i(\omega)}{n} \ln p_i\right] - H\right| \leqslant -\sum_{i=1}^{r} \left|\frac{\nu_i(\omega)}{n} - p_i\right| \times \ln p_i \leqslant -\varepsilon \sum_{i=1}^{r} \ln p_i.$$

由此可见, 典型轨道的概率接近 e^{-nH}, 因为由于大数定律当 n 充分大时, 典型轨道的条数 "几乎" 穷尽 Ω 中的所有点, 而 Ω 中轨道的条数应该为量级 e^{nH}. 将以上的讨论归纳为下面的定理.

定理 (麦克米兰 [B. McMillan]) 设 $p_i > 0 (i = 1, \cdots, r), 0 < \varepsilon < 1$,则存在一自然数 $n_0 = n_0(\varepsilon; p_1, \cdots, p_r)$, 使对于一切 $n > n_0$:

a) $e^{n(H-\varepsilon)} \leqslant N(C(n, \varepsilon_1)) \leqslant e^{n(H+\varepsilon)}$,

b) $e^{-n(H+\varepsilon)} \leqslant p(\omega) \leqslant e^{-n(H-\varepsilon)}, \omega \in C(n, \varepsilon_1)$,

c) $\mathbf{P}(C(n, \varepsilon_1)) = \sum\limits_{\omega \in C(n,\varepsilon_1)} p(\omega) \to 1 (n \to \infty)$,

其中

$$\varepsilon_1 = \min \left[\varepsilon, \varepsilon \left(-2 \sum_{k=1}^{r} \ln p_k \right)^{-1} \right].$$

证明 命题 c) 由大数定律得出. 对于其余命题注意到, 如果 $\omega \in C(n, \varepsilon_1)$, 则

$$np_k - \varepsilon_1 n < \nu_k(\omega) < np_k + \varepsilon_1 n \quad (k = 1, \cdots, r),$$

因此

$$\begin{aligned} p(\omega) = \exp\left\{-\sum \nu_k \ln p_k\right\} &< \exp\left\{-n \sum p_k \ln p_k - \varepsilon_1 n \sum \ln p_k\right\} \\ &\leqslant \exp\left\{-n\left(H - \frac{\varepsilon}{2}\right)\right\}. \end{aligned}$$

类似地, 有

$$p(\omega) > \exp\left\{-n\left(H + \frac{\varepsilon}{2}\right)\right\}.$$

从而, 命题 b) 得证.

最后, 由于

$$\mathbf{P}(C(n, \varepsilon_1)) \geqslant N(C(n, \varepsilon_1)) \times \min_{\omega \in C(n,\varepsilon_1)} p(\omega),$$

则

$$N(C(n, \varepsilon_1)) \leqslant \frac{\mathbf{P}(C(n, \varepsilon_1))}{\min\limits_{\omega \in C(n,\varepsilon_1)} p(\omega)} < \frac{1}{e^{-n\left(H+\frac{\varepsilon}{2}\right)}} = e^{n\left(H+\frac{\varepsilon}{2}\right)};$$

类似地, 有

$$N(C(n, \varepsilon_1)) \geqslant \frac{\mathbf{P}(C(n, \varepsilon_1))}{\max\limits_{\omega \in C(n,\varepsilon_1)} p(\omega)} > \mathbf{P}(C(n, \varepsilon_1)) e^{n\left(H-\frac{\varepsilon}{2}\right)}.$$

由于 $\mathbf{P}(C(n, \varepsilon_1)) \to 1 (n \to \infty)$, 可见存在 n_1, 使对于 $n > n_1$, 有 $\mathbf{P}(C(n, \varepsilon_1)) > 1 - \varepsilon$, 故

$$N(C(n, \varepsilon_1)) \geqslant (1 - \varepsilon) \exp\left\{n\left(H - \frac{\varepsilon}{2}\right)\right\} = \exp\left\{n(H - \varepsilon) + \left[\frac{n\varepsilon}{2} + \ln(1 - \varepsilon)\right]\right\}.$$

假设 n_2 满足: 对于 $n > n_2$, 有

$$\frac{n\varepsilon}{2} + \ln(1-\varepsilon) > 0.$$

于是, 对于 $n \geqslant n_0 = \max\{n_1, n_2\}$, 有

$$N(C(n,\varepsilon_1)) \geqslant \mathrm{e}^{n(H-\varepsilon)}.$$ □

5. 用概率方法证明维尔斯特拉斯定理 利用伯努利概型的大数定律, 可以给著名的维尔斯特拉斯 (K. T. W. Wierstrass) 定理 "用多项式逼近连续函数" 以简单而雅致的证明.

设 $f(p)$ 是线段 $[0,1]$ 上的连续函数. 引进多项式

$$B_n(p) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \mathrm{C}_n^k p^k (1-p)^{n-k}, \quad 0 \leqslant p \leqslant 1, n \geqslant 0. \tag{15}$$

该多项式称做伯恩斯坦 (С. Н. Бернштейн) 多项式①, 用提供维尔斯特拉斯定理的此证明的作者命名.

如果 $\xi_1, \cdots, \xi_n$ 是独立伯努利随机变量序列, 且 $\mathbf{P}\{\xi_i = 1\} = p, \mathbf{P}\{\xi_i = 1\} = q$, 设 $S_n = \xi_1 + \cdots + \xi_n$, 则

$$\mathbf{E} f(S_n/n) = B_n(p).$$

由于在闭区间 $[0,1]$ 上的连续函数 $f = f(p)$ 一致连续, 可见对于任意 $\varepsilon > 0$ 存在 $\delta > 0$, 使当 $|x-y| \leqslant \delta$ 时 $|f(x) - f(y)| \leqslant \varepsilon$. 显然, 这样的函数有界: $|f(x)| \leqslant M < \infty$.

因此由不等式 (5), 可见

$$\begin{aligned}
|f(p) - B_n(p)| &= \left| \sum_{k=0}^{n} \left[f(p) - f\left(\frac{k}{n}\right) \right] \mathrm{C}_n^k p^k q^{n-k} \right| \\
&\leqslant \sum_{\left\{k: \left|\frac{k}{n} - p\right| \leqslant \delta\right\}} \left| f(p) - f\left(\frac{k}{n}\right) \right| \mathrm{C}_n^k p^k q^{n-k} \\
&\quad + \sum_{\left\{k: \left|\frac{k}{n} - p\right| > \delta\right\}} \left| f(p) - f\left(\frac{k}{n}\right) \right| \mathrm{C}_n^k p^k q^{n-k} \\
&\leqslant \varepsilon + 2M \sum_{\left\{k: \left|\frac{k}{n} - p\right| > \delta\right\}} \mathrm{C}_n^k p^k q^{n-k} \leqslant \varepsilon + \frac{2M}{4n\delta^2} = \varepsilon + \frac{M}{2n\delta^2}.
\end{aligned}$$

由此可见, 对于伯恩斯坦多项式 (15),

$$\lim_{n\to\infty} \max_{0 \leqslant p \leqslant 1} |f(p) - B_n(p)| = 0,$$

这正是维尔斯特拉斯定理的结论.

① С. Н. 伯恩斯坦 (С. Н. Бернштейн , 1880 — 1968, 乌兹别克斯坦统计学家). —— 译者

6. 练习题

1. 设随机变量 ξ 和 η 的相关系数为 ρ. 证明切比雪夫不等式二维类似:

$$\mathbf{P}\{|\xi-\mathbf{E}\xi|\geqslant\varepsilon\sqrt{\mathbf{D}\xi}\ \text{或}\ |\eta-\mathbf{E}\eta|\geqslant\varepsilon\sqrt{\mathbf{D}\eta}\}\leqslant\frac{1+\sqrt{1-\rho^2}}{\varepsilon^2}.$$

(提示. 利用 §4 中练习题 8 的结果.)

2. 设 $f=f(x)$ 为非负偶函数, 且当 $x>0$ 时非减. 设 $\xi=\xi(\omega)$ 是非负随机变量, 且 $|\xi(\omega)|\leqslant C$, 证明对于任意 $\varepsilon>0$, 有

$$\mathbf{P}\{|\xi|\geqslant\varepsilon\}\geqslant\frac{\mathbf{E}f(\xi)-f(\varepsilon)}{f(C)}.$$

特别, 对于 $f(x)=x^2$,

$$\frac{\mathbf{E}\xi^2-\varepsilon^2}{C^2}\leqslant\mathbf{P}\{|\xi-\mathbf{E}\xi|\geqslant\varepsilon\}\leqslant\frac{\mathbf{D}\xi}{\varepsilon^2}.$$

3. 设 $\xi_1,\cdots,\xi_n$ 是独立随机变量序列, $\mathbf{D}\xi_i\leqslant C$, 证明

$$\mathbf{P}\left\{\left|\frac{\xi_1+\cdots+\xi_n}{n}-\frac{\mathbf{E}(\xi_1+\cdots+\xi_n)}{n}\right|\geqslant\varepsilon\right\}\leqslant\frac{C}{n\varepsilon^2}. \tag{16}$$

(与关系式 (8) 有同样的补充说明, 由不等式 (16), 可见在比伯努利概型更一般的情形下, 大数定律依然成立.)

4. 设 $\xi_1,\cdots,\xi_n$ 是独立伯努利随机变量, 且 $\mathbf{P}\{\xi_i=1\}=p>0,\mathbf{P}\{\xi_i=-1\}=q(p+q=1)$, 证明有如下伯恩斯坦估计: 对于任意 $a>0$, 有

$$\mathbf{P}\left\{\left|\frac{S_n}{n}-(2p-1)\right|\geqslant\varepsilon\right\}\leqslant 2\mathrm{e}^{-a\varepsilon^2 n},$$

其中 $S_n=\xi_1+\cdots+\xi_n,\varepsilon>0$.

5. 设 ξ 是非负随机变量, 而 $a>0$. 求概率 $\mathbf{P}\{\xi\geqslant a\}$ 的上确界, 假如已知

(1) $\mathbf{E}\xi=20$;

(2) $\mathbf{E}\xi=20,\mathbf{D}\xi=25$;

(3) $\mathbf{E}\xi=20,\mathbf{D}\xi=25$, 且 ξ 关于数学期望对称.

§6. 伯努利概型 II. 极限定理 (棣莫弗 – 拉普拉斯局部定理、泊松定理)

1. 棣莫弗 – 拉普拉斯局部定理 像上一节一样, 考虑

$$S_n=\xi_1+\cdots+\xi_n.$$

那么

$$\mathbf{E}\frac{S_n}{n}=p, \tag{1}$$

而由 §4, (14) 式

$$\mathbf{E}\left(\frac{S_n}{n}-p\right)^2=\frac{pq}{n}. \tag{2}$$

由 (1) 式可见, $S_n/n \sim p$, 其中关于等价符号 "$\sim$" 曾以概率

$$\mathbf{P}\left\{\left|\frac{S_n}{n}-p\right|\geqslant\varepsilon\right\}$$

的估计的形式, 在大数定律中得到确切的解释. 自然想到, 由 "关系式" (2) 亦可给

$$\left|\frac{S_n}{n}-p\right|\sim\sqrt{\frac{pq}{n}} \tag{3}$$

以确切的概率意义. 例如, 考虑形如

$$\mathbf{P}\left\{\left|\frac{S_n}{n}-p\right|\leqslant x\sqrt{\frac{pq}{n}}\right\},\quad x\in\mathbb{R}^1$$

的概率, 或 (由于 $\mathbf{E}S_n=np, \mathbf{D}S_n=npq$) 考虑概率

$$\mathbf{P}\left\{\left|\frac{S_n-\mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}}\right|\leqslant x\right\}.$$

如果对 $n\geqslant 1$, 仍记

$$P_n(k)=\mathrm{C}_n^k p^k q^{n-k},\quad 0\leqslant k\leqslant n,$$

则概率

$$\mathbf{P}\left\{\left|\frac{S_n-\mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}}\right|\leqslant x\right\}=\sum_{\left\{k:\frac{|k-np|}{\sqrt{npq}}\leqslant x\right\}}P_n(k). \tag{4}$$

现在提出问题: 当 $n\to\infty$ 时求概率 $P_n(k)$ 及其和

$$\sum_{\left\{k:\frac{|k-np|}{\sqrt{npq}}\leqslant x\right\}}P_n(k)$$

满足的便于应用的渐近公式.

下面的定理, 对于既满足 $|k-np|=O(\sqrt{npq})$ 又满足 $|k-np|=o(npq)^{2/3}$ 的 k 值, 给出了答案.

局部极限定理 设 $0<p<1$, 则对满足 $|k-np|=o(npq)^{2/3}$ 的所有 k, 一致有

$$P_n(k)\sim\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}}, \tag{5}$$

即当 $n\to\infty$ 时

$$\sup_{\{k:|k-np|\leqslant\varphi(n)\}}\left|\frac{P_n(k)}{\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}}}-1\right|\to 0, \tag{6}$$

其中 $\varphi(n)$ 是任意满足 $\varphi(n) = o(npq)^{2/3}$ 的非负函数.

证明　对于证明, 主要用到斯特林 (J. Stering) 公式 (§6, (2) 式):

$$n! = \sqrt{2\pi n}\mathrm{e}^{-n}n^n(1+R(n)),$$

其中当 $n\to\infty$ 时 $R(n)\to 0$.

根据斯特林公式, 若 $n\to\infty, k\to\infty, n-k\to\infty$, 则

$$\begin{aligned}
\mathrm{C}_n^k &= \frac{n!}{k!(n-k)!}\\
&= \frac{\sqrt{2\pi n}\mathrm{e}^{-n}n^n(1+R(n))}{\sqrt{2\pi k\times 2\pi(n-k)}\mathrm{e}^{-k}k^k(n-k)^{n-k}\mathrm{e}^{-(n-k)}(1+R(k))(1+R(n-k))}\\
&= \frac{1}{\sqrt{2\pi n\dfrac{k}{n}\left(1-\dfrac{k}{n}\right)}}\times\frac{1+\varepsilon(n,k,n-k)}{\left(\dfrac{k}{n}\right)^k\left(1-\dfrac{k}{n}\right)^{n-k}},
\end{aligned}$$

其中当 $n\to\infty, k\to\infty, n-k\to\infty$ 时, 显然 $\varepsilon(n,k,n-k)\to 0$.

因此

$$P_n(k) = \mathrm{C}_n^k p^k q^{n-k} = \frac{1}{\sqrt{2\pi n\dfrac{k}{n}\left(1-\dfrac{k}{n}\right)}}\times\frac{p^k(1-p)^{n-k}(1+\varepsilon)}{\left(\dfrac{k}{n}\right)^k\left(1-\dfrac{k}{n}\right)^{n-k}}.$$

记 $\widehat{p} = k/n$, 则

$$\begin{aligned}
P_n(k) &= \frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\left(\frac{p}{\widehat{p}}\right)^k\left(\frac{1-p}{1-\widehat{p}}\right)^{n-k}(1+\varepsilon)\\
&= \frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\exp\left\{k\ln\frac{p}{\widehat{p}}+(n-k)\ln\frac{1-p}{1-\widehat{p}}\right\}\times(1+\varepsilon)\\
&= \frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\exp\left\{n\left[\frac{k}{n}\ln\frac{p}{\widehat{p}}+\left(1-\frac{k}{n}\right)\ln\frac{1-p}{1-\widehat{p}}\right]\right\}\times(1+\varepsilon)\\
&= \frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\exp\{-nH(\widehat{p})\}\times(1+\varepsilon),
\end{aligned}$$

其中

$$H(x) = x\ln\frac{x}{p}+(1-x)\ln\frac{1-x}{1-p}.$$

所考虑的 k 满足 $|k-np| = o(npq)^{2/3}$, 故当 $n\to\infty$ 时 $p-\widehat{p}\to 0$.

由于对 $0<x<1$

$$\begin{aligned}
H'(x) &= \ln\frac{x}{p}-\ln\frac{1-x}{1-p},\\
H''(x) &= \frac{1}{x}+\frac{1}{1-x},\\
H'''(x) &= -\frac{1}{x^2}+\frac{1}{(1-x)^2}.
\end{aligned}$$

因此, 若将 $H(\widehat{p})$ 表示为 $H(p+(\widehat{p}-p))$, 并利用泰勒 (B. Taylor) 公式, 则当 n 充分大时, 有

$$\begin{aligned}H(\widehat{p}) &= H(p)+H'(p)(\widehat{p}-p)+\frac{1}{2}H''(p)(\widehat{p}-p)^2+O(|\widehat{p}-p|^3)\\ &= \frac{1}{2}\left(\frac{1}{p}+\frac{1}{q}\right)(\widehat{p}-p)^2+O(|\widehat{p}-p|^3).\end{aligned}$$

从而

$$P_n(k)=\frac{1}{\sqrt{2\pi n\widehat{p}(1-\widehat{p})}}\exp\left\{-\frac{n}{2pq}(\widehat{p}-p)^2+nO(|\widehat{p}-p|^3)\right\}\times(1+\varepsilon.)$$

注意到

$$\frac{n}{2pq}(\widehat{p}-p)^2=\frac{n}{2pq}\left(\frac{k}{n}-p\right)^2=\frac{(k-np)^2}{2npq}.$$

因此

$$P_n(k)=\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}}(1+\varepsilon'(n,k,n-k)),$$

其中

$$1+\varepsilon'(n,k,n-k)=[1+\varepsilon(n,k,n-k)]\mathrm{e}^{nO(|\widehat{p}-p|^3)}\sqrt{\frac{p(1-p)}{\widehat{p}(1-\widehat{p})}}.$$

易见

$$\sup|\varepsilon'(n,k,n-k)|\to 0,\quad n\to\infty,$$

其中 sup 对满足

$$|k-np|\leqslant\varphi(n),\quad \varphi(n)=o(npq)^{2/3}$$

的 k 来求. □

系 *局部极限定理的结论, 可以表述为如下等价的形式: 对于一切 $x\in\mathbb{R}^1$, 若 $x=o(npq)^{1/6}$, 而 $np+x\sqrt{npq}$ 是集合 $\{0,1,\cdots,n\}$ 中的整数, 则*

$$P_n(np+x\sqrt{npq})\sim\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{x^2}{2}},\tag{7}$$

即当 $n\to\infty$ 时, 有

$$\sup_{\{x:|x|\leqslant\psi(n)\}}\left|\frac{P_n(np+x\sqrt{npq})}{\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{x^2}{2}}}-1\right|\to 0\tag{8}$$

其中 $\psi(n)=o(npq)^{1/6}$.

注意到关于 §5 中 (8) 式的说明, 可以用概率的语言将上面得到的结果表述为:

$$\mathbf{P}\{S_n=k\}\sim\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{(k-np)^2}{2npq}},|k-np|=o(npq)^{2/3},\tag{9}$$

$$\mathbf{P}\left\{\frac{S_n-np}{\sqrt{npq}}=x\right\}\sim\frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-\frac{x^2}{2}},x=o(npq)^{1/6}.\tag{10}$$

(假设在 (10) 式中 $np+x\sqrt{npq}$ 取 $0,1,\cdots,n$ 为值.)

假如设

$$t_k=\frac{k-np}{\sqrt{npq}},\quad \Delta t_k=t_{k+1}-t_k=\frac{1}{\sqrt{npq}},$$

则 (10) 式具有如下形式:

$$\mathbf{P}\left\{\frac{S_n-\mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}}=t_k\right\}\sim\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}},\quad t_k=o(npq)^{1/6}. \tag{11}$$

显然当 $n\to\infty$ 时

$$\Delta t_k=\frac{1}{\sqrt{npq}}\to 0,$$

而点 t_k 的集合 $\{t_k\}$ "充满" 整个数轴. 因此, 自然想到由 (11) 式可以得到积分公式:

$$\mathbf{P}\left\{a<\frac{S_n-np}{\sqrt{npq}}\leqslant b\right\}\sim\frac{1}{\sqrt{2\pi}}\int_a^b\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x,\quad -\infty<a\leqslant b<\infty.$$

下面将给出确切的表述.

2. 棣莫弗 – 拉普拉斯积分定理　对于 $-\infty<a\leqslant b<\infty$, 设

$$P_n(a,b]=\sum_{a<x\leqslant b}P_n(np+x\sqrt{npq}),$$

其中对一切使 $np+x\sqrt{npq}$ 为整数的 x 求和.

由局部定理可见 (亦见 (11) 式), 对于由 $k=np+t_k\sqrt{npq}$ 决定且满足条件 $|t_k|\leqslant T<\infty$ 的 t_k, 有

$$P_n(np+t_k\sqrt{npq})=\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}[1+\varepsilon(t_k,n)], \tag{12}$$

其中

$$\sup_{|t_k|\leqslant T}|\varepsilon(t_k,n)|\to 0, n\to\infty. \tag{13}$$

从而, 对于固定的 a,b ($-T\leqslant a\leqslant b\leqslant T$, 而 $T<\infty$),

$$\begin{aligned}\sum_{a<t_k\leqslant b}P_n(np+t_k\sqrt{npq})&=\sum_{a<t_k\leqslant b}\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}+\sum_{a<t_k\leqslant b}\varepsilon(t_k,n)\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}\\&=\frac{1}{\sqrt{2\pi}}\int_a^b\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x+R_n^{(1)}(a,b)+R_n^{(2)}(a,b).\end{aligned} \tag{14}$$

其中

$$R_n^{(1)}(a,b)=\sum_{a<t_k\leqslant b}\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}-\frac{1}{\sqrt{2\pi}}\int_a^b\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x,$$

$$R_n^{(2)}(a,b)=\sum_{a<t_k\leqslant b}\varepsilon(t_k,n)\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}.$$

由熟知的积分和的性质,

$$\sup_{-T\leqslant a\leqslant b\leqslant T}|R_n^{(1)}(a,b)|\to 0,\quad n\to\infty. \tag{15}$$

同样易见

$$\sup_{-T\leqslant a\leqslant b\leqslant T}|R_n^{(2)}(a,b)|\leqslant \sup_{|t_k|\leqslant T}|\varepsilon(t_k,n)|\sum_{|t_k|\leqslant T}\frac{\Delta t_k}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t_k^2}{2}}$$

$$\leqslant \sup_{|t_k|\leqslant T}|\varepsilon(t_k,n)|\times\left[\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x+\sup_{-T\leqslant a\leqslant b\leqslant T}|R_n^{(1)}(a,b)|\right]\to 0. \tag{16}$$

其中右侧收敛于 0, 是因为 (15) 式以及数学分析中熟知的事实

$$\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\leqslant\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x=1. \tag{17}$$

记

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t,$$

则由 (14) $\sim$ (16) 式可见

$$\sup_{-T\leqslant a\leqslant b\leqslant T}|P_n(a,b]-[\Phi(b)-\Phi(a)]|\to 0, n\to\infty. \tag{18}$$

现在证明, 该式不仅对于有限 T 成立, 而且对于 $T=\infty$ 也成立. 由于 (17) 式, 对于任意给定的 $\varepsilon>0$, 存在有限 $T=T(\varepsilon)$, 使

$$\frac{1}{\sqrt{2\pi}}\int_{-T}^{T}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x>1-\frac{\varepsilon}{4}. \tag{19}$$

根据 (18) 式, 对于任意 $\varepsilon>0$, 存在 N, 使对于一切 $n>N$ 和 $T=T(\varepsilon)$, 有

$$\sup_{-T\leqslant a\leqslant b\leqslant T}|P_n(a,b]-[\Phi(b)-\Phi(a)]|<\frac{\varepsilon}{4}. \tag{20}$$

由此和 (19) 式, 可见

$$P_n(-T,T]>1-\frac{\varepsilon}{2},$$

因此

$$P_n(-\infty,-T]+P_n(T,\infty)\leqslant\frac{\varepsilon}{2},$$

其中

$$P_n(-\infty,T]=\lim_{S\downarrow-\infty}P_n(S,T], P_n(T,\infty)=\lim_{S\uparrow\infty}P_n(T,S].$$

这样, 对于任意 $-\infty \leqslant -T \leqslant a \leqslant b \leqslant T \leqslant \infty$, 有

$$\begin{aligned}&\left|P_n(a,b]-\frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right| \leqslant \left|P_n(-T,T]-\frac{1}{\sqrt{2\pi}}\int_{-T}^{T} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right| \\ &\qquad + \left|P_n(a,-T]-\frac{1}{\sqrt{2\pi}}\int_a^{-T} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right| + \left|P_n(T,b]-\frac{1}{\sqrt{2\pi}}\int_T^b \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right| \\ &\leqslant \frac{\varepsilon}{4}+P_n(-\infty,-T]+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-T} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x + P_n(T,\infty)+\frac{1}{\sqrt{2\pi}}\int_T^{\infty} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x \\ &\leqslant \frac{\varepsilon}{4}+\frac{\varepsilon}{2}+\frac{\varepsilon}{8}+\frac{\varepsilon}{8}=\varepsilon.\end{aligned}$$

注意到 (18) 式, 由此容易证明 $P_n(a,b]$ 关于 $-\infty \leqslant a \leqslant b \leqslant \infty$ 一致趋向 $\Phi(b)-\Phi(a)$. 于是, 证明了下面的定理.

棣莫弗 - 拉普拉斯积分定理　设 $0<p<1$,

$$P_n(k)=\mathrm{C}_n^k p^k q^{n-k},\ P_n(a,b]=\sum_{a<x\leqslant b} P_n(np+x\sqrt{npq}).$$

那么

$$\sup_{-\infty\leqslant a<b\leqslant\infty}\left|P_n(a,b]-\frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right| \to 0,\quad n\to\infty. \tag{21}$$

精确到 §5 中 (8) 式所指, 可以将 (21) 式的结果用概率的语言表示为:

$$\sup_{-\infty\leqslant a<b\leqslant\infty}\left|\mathbf{P}\left\{a<\frac{S_n-\mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}}\leqslant b\right\}-\frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right| \to 0,\quad n\to\infty.$$

由此可见, 对于任意 $-\infty \leqslant A < B \leqslant \infty$, 有

$$\mathbf{P}\{A<S_n\leqslant B\}-\left[\Phi\left(\frac{B-np}{\sqrt{npq}}\right)-\Phi\left(\frac{A-np}{\sqrt{npq}}\right)\right]\to 0,\quad n\to\infty. \tag{22}$$

例　将规则的色子掷 12 000 次. 问 “6 点” 出现的次数属于区间 (1800, 2100] 的概率 P 如何?

所求概率等于

$$P=\sum_{1800<k\leqslant 2100}\mathrm{C}_{12\,000}^k\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k}.$$

显然, 用 “手算” 精确地算出该和的值是相当困难的. 假如利用积分定理, 则求得此概率 P 大致等于 $(n=12\,000, p=1/6, A=1800, B=2100)$:

$$\Phi\left(\frac{2100-2000}{\sqrt{12\,000\times\frac{1}{6}\times\frac{5}{6}}}\right)-\Phi\left(\frac{1800-2000}{\sqrt{12\,000\times\frac{1}{6}\times\frac{5}{6}}}\right)$$
$$=\Phi(\sqrt{6})-\Phi(-2\sqrt{6})\approx\Phi(2.449)-\Phi(-4.898)\approx 0.992,$$

其中 $\Phi(2.449)$ 和 $\Phi(-4.898)$ 的值, 由正态分布函数的 $\Phi(x)$ 数值表查出 (参见第 6 小节).

3. 二项概率的正态逼近 把二项概率 $P_n(np+x\sqrt{npq})$ (假设只考虑使 $np+x\sqrt{npq}$ 为整数的 x) 标在图上 (图 9).

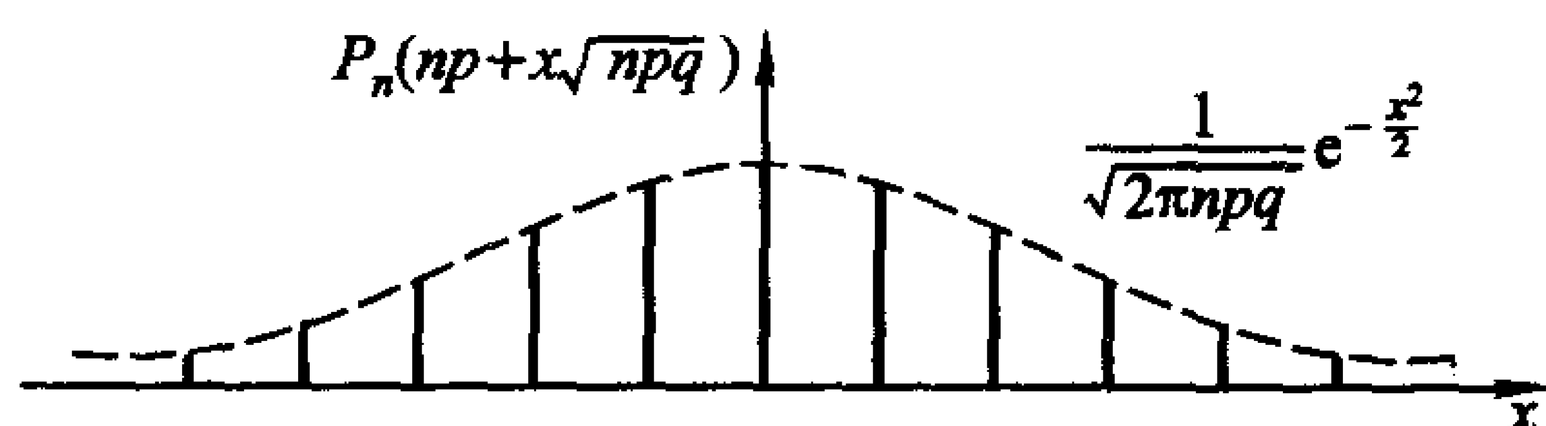


图 9

那么, 局部定理表明, 对于 $x=o(npq)^{1/6}$, 概率 $P_n(np+x\sqrt{npq})$ 的值较好地 "位于" 正态密度曲线上:

$$\frac{1}{\sqrt{2\pi npq}}e^{-\frac{x^2}{2}}.$$

由积分定理知, 概率

$$\begin{aligned}P_n(a,b]&=\mathbf{P}\{a\sqrt{npq}<S_n-np\leqslant b\sqrt{npq}\}\\&=\mathbf{P}\{np+a\sqrt{npq}<S_n\leqslant np+b\sqrt{npq}\}\end{aligned}$$

的值可以较好地由积分

$$\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\mathrm{d}x$$

逼近.

记

$$F_n(x)=P_n(-\infty,x]\left(=\mathbf{P}\left\{\frac{S_n-np}{\sqrt{npq}}\leqslant x\right\}\right).$$

那么, 由 (21) 式可见

$$\sup_{-\infty\leqslant x\leqslant\infty}|F_n(x)-\Phi(x)|\to 0, n\to\infty. \tag{23}$$

重要的是在 (21) 和 (23) 式中, 随着 n 的增长趋向 0 的速度如何. 这里引用的结果是贝里 - 埃森 (A. C. Berry-C. G. Esseen) 定理的特殊情形 (第三章 §11):

$$\sup_{-\infty\leqslant x\leqslant\infty}|F_n(x)-\Phi(x)|\leqslant\frac{p^2+q^2}{\sqrt{npq}}. \tag{24}$$

要特别强调, 估计 $1/\sqrt{npq}$ 的量级不能再提高了. 这指的是, 当 p 的值接近 0 或 1 时, 甚至对于充分大的 n, 用函数 $\Phi(x)$ 逼近 $F_n(x)$ 的效果可能不佳. 因此产生一个问题, 当 p 或 q 的值较小时, 能否为我们关心的概率, 找到比局部和积分定理给出的正态分布更好的逼近. 为此我们指出, 当 $p=1/2$ 时二项分布 $\{P_n(k)\}$ 具有对称的形状 (图 10 左边的图). 不过, 当 p 较小时, 二项分布的形状是非对称的 (图 10 的右边的图), 不能指望用正态逼近有好结果.

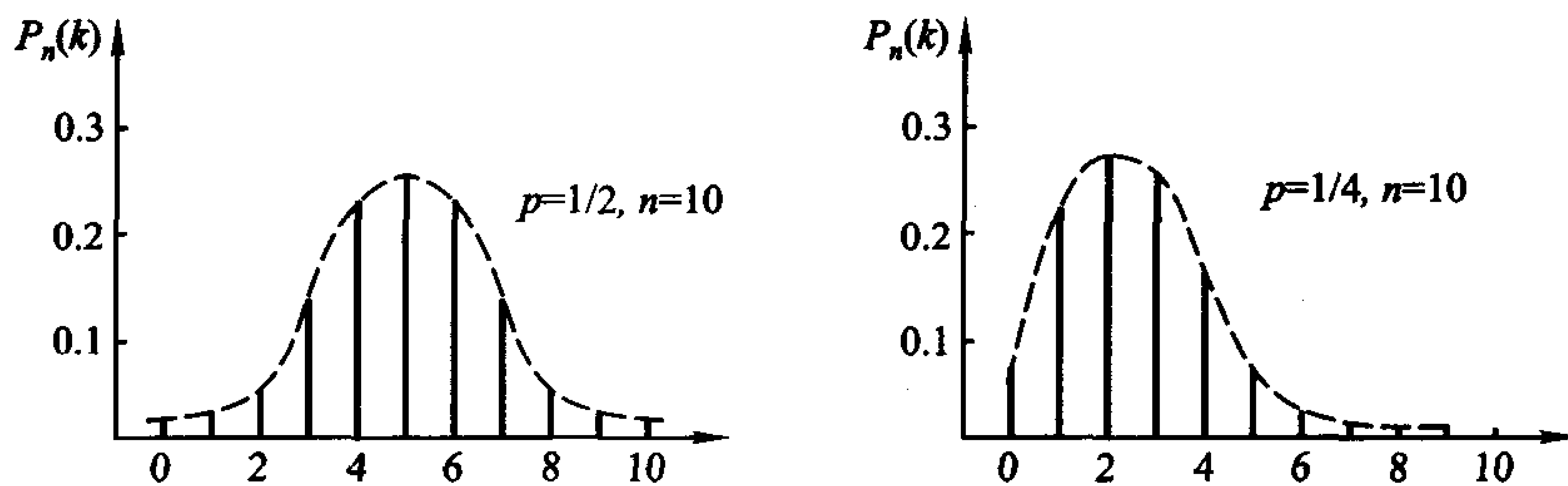


图 10

4. 泊松定理 结果表明, 对于较小的 p 值, 概率的所谓泊松分布可以很好地逼近二项分布的概率 $P_n(k)$.

设

$$P_n(k) = \begin{cases} C_n^k p^k q^{n-k}, & 若\ k = 0, 1, \cdots, n, \\ 0, & 若\ k = n+1, n+2, \cdots, \end{cases}$$

且假设 p 是 n 的函数 $p = p(n)$.

泊松定理 设 $p(n) \to 0, n \to \infty$, 且 $np(n) \to \lambda$, 其中 $\lambda > 0$. 那么, 对于任意 $k = 0, 1, \cdots$, 有

$$P_n(k) \to \pi_k, \quad n \to \infty, \tag{25}$$

其中

$$\pi_k = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \cdots. \tag{26}$$

证明 证明十分简单. 由于根据条件

$$p(n) = \frac{\lambda}{n} + o\left(\frac{1}{n}\right),$$

可见对于固定的 $k = 0, 1, \cdots$ 和充分大的 n,

$$\begin{aligned} P_n(k) &= C_n^k p^k q^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!}\left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k \times \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k}. \end{aligned}$$

由

$$\begin{aligned} & n(n-1)\cdots(n-k+1)\left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k \\ = & \frac{n(n-1)\cdots(n-k+1)}{n^k}[\lambda + o(1)]^k \to \lambda^k, n \to \infty. \end{aligned}$$

和

$$\left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k} \to e^{-\lambda}, n \to \infty,$$

立即得 (25) 式. □

数组 $\{\pi_k, k=0,1\cdots\}$, 满足

$$\pi_k > 0, \sum_{k=0}^{\infty} \pi_k = 1,$$

因此可以作为概率分布, 称做*泊松分布*. 注意, 以上所讨论的概率分布都集中在有限个点上. 泊松分布, 是我们第一次遇到的集中在可数个点上的 (离散型) 分布的例子.

下面引进的 (Ю. В. 普罗霍洛夫的) 结果, 给出了当 $n\to\infty$ 时概率 $P_n(k)$ 收敛于 π_k 的速度: 如果 $np(n)=\lambda$, 则*)

$$\sum_{k=0}^{\infty} |P_n(k) - \pi_k| \leqslant \frac{2\lambda}{n} \times \min(2, \lambda). \tag{27}$$

5. 棣莫弗 – 拉普拉斯定理与大数定律 我们再回到棣莫弗 – 拉普拉斯极限定理. (在 §5 对 (8) 式说明的前提下) 说明如何由棣莫弗 – 拉普拉斯极限定理, 得出大数定律. 因为

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} = \mathbf{P}\left\{\left|\frac{S_n - np}{\sqrt{npq}}\right| \leqslant \varepsilon\sqrt{\frac{n}{pq}}\right\},$$

所以由 (21) 式可见, 对于 $\varepsilon > 0$, 当 $n\to\infty$ 时

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} - \frac{1}{\sqrt{2\pi}}\int_{-\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x \to 0, \tag{28}$$

因此

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \to 1, \quad n\to\infty,$$

此即大数定律的结论.

由 (28) 式, 可见

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \sim \frac{1}{\sqrt{2\pi}}\int_{-\varepsilon\sqrt{n/pq}}^{\varepsilon\sqrt{n/pq}} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x, \quad n\to\infty, \tag{29}$$

然而, 切比雪夫不等式只能给出下面的估计

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \frac{pq}{n\varepsilon^2}.$$

在 §5 第 3 小节关于为使不等式

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \alpha$$

---

*) 较弱结果的证明, 将在第三章 §12 介绍.

成立所需要的观测次数, 由切比雪夫不等式得到下面的估计

$$n \geqslant \left[\frac{1}{4\varepsilon^2\alpha}\right] (= n_1(\alpha)),$$

其中 $[x]$ 是 x 的整数部分. 例如, 对于 $\varepsilon = 0.02, \alpha = 0.05$, 需要 12 500 次观测. 现在, 利用等价关系式 (29) 解决了同一问题.

我们由关系式

$$\frac{1}{\sqrt{2\pi}} \int_{-k(\alpha)}^{k(\alpha)} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = 1 - \alpha$$

求 $k(\alpha)$. 由于

$$\varepsilon\sqrt{\frac{n}{pq}} \geqslant 2\varepsilon\sqrt{n},$$

并由不等式

$$2\varepsilon\sqrt{n} \geqslant k(\alpha), \tag{30}$$

求出 (最小整数), 得

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \alpha. \tag{31}$$

由 (30) 式, 可见 $n = n_2(\alpha)$, 其中

$$n_2(\alpha) = \left[\frac{k^2(\alpha)}{4\varepsilon^2}\right]$$

可以保证 (31) 式成立, 其逼近精度容易由 (24) 式得到.

取 $\varepsilon = 0.02, \alpha = 0.05$, 可见只需要 2 500 次观测, 而不是切比雪夫不等式要求的 12 500 次. 下面对于一些 α 值, 列举相应的 $k(\alpha)$ 值:

α:	0.50	0.317 3	0.10	0.05	0.045 4	0.01	0.002 7
$k(\alpha)$:	0.675	1.000	1.645	1.960	2.000	2.576	3.000

6. 正态分布 前面在棣莫弗 - 拉普拉斯积分定理里, 引进的函数

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t, \tag{32}$$

在概率论里起非常重要的作用, 称做*正态分布函数*或*高斯分布函数*. 函数

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}^1,$$

称做*正态密度*或*高斯密度*.

我们已经见到过在*有限或可数*点集上的 (离散型) 分布. 正态分布属于概率论中另外一种非常重要类型的分布 —— 连续型分布. 正态分布之所以非常重要, 首先是

因为在相当一般的条件下，大量独立随机变量 (未必是伯努利变量) 之和的分布，可以很好地用正态分布来逼近 (第三章 §4). 我们现在讨论函数 $\varphi(x)$ 和 $\Phi(x)$ 的一些简单性质，而图 11 和图 12 分别是 $\varphi(x)$ 和 $\Phi(x)$ 的图形.

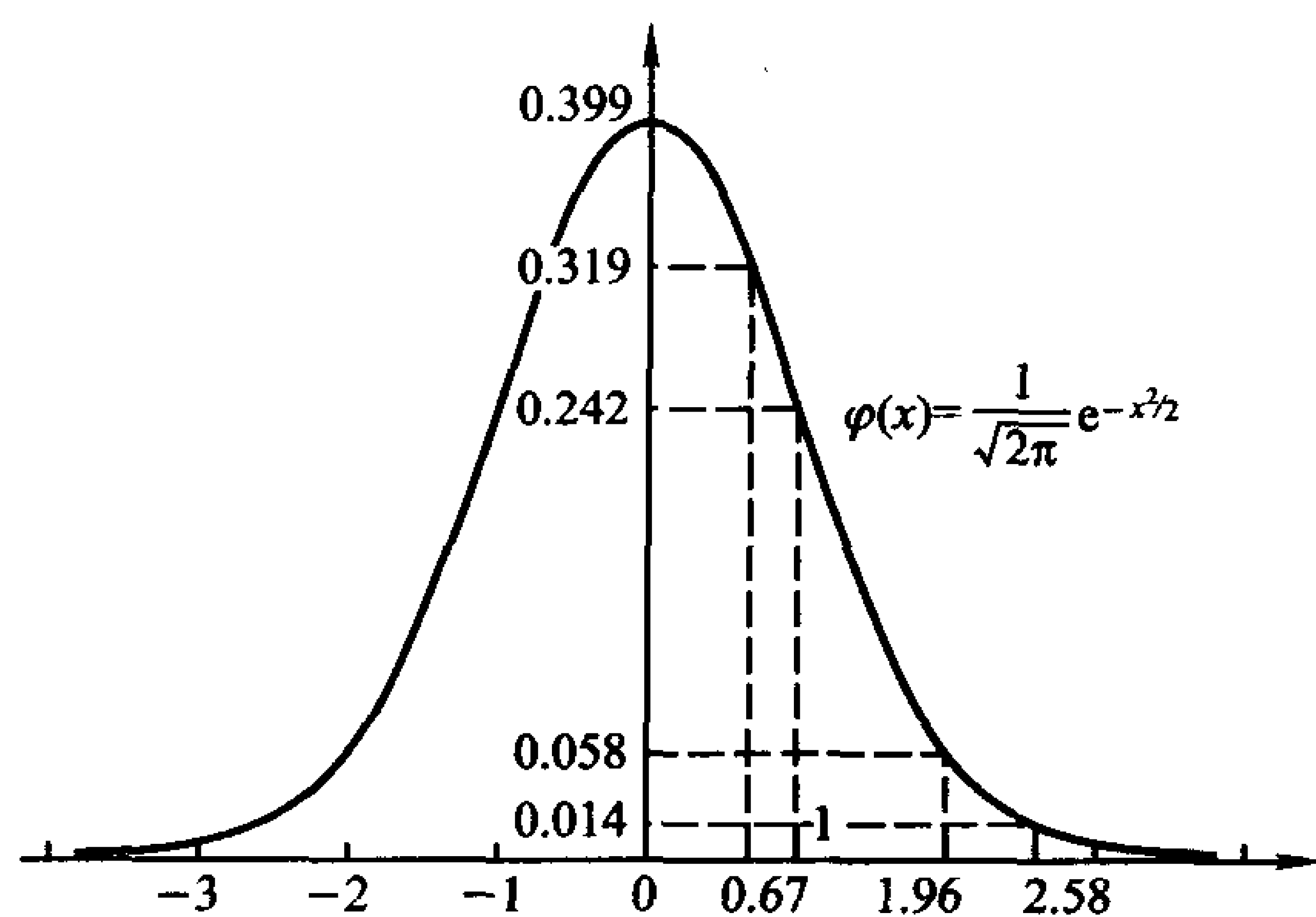


图 11 正态分布密度 $\varphi(x)$ 图

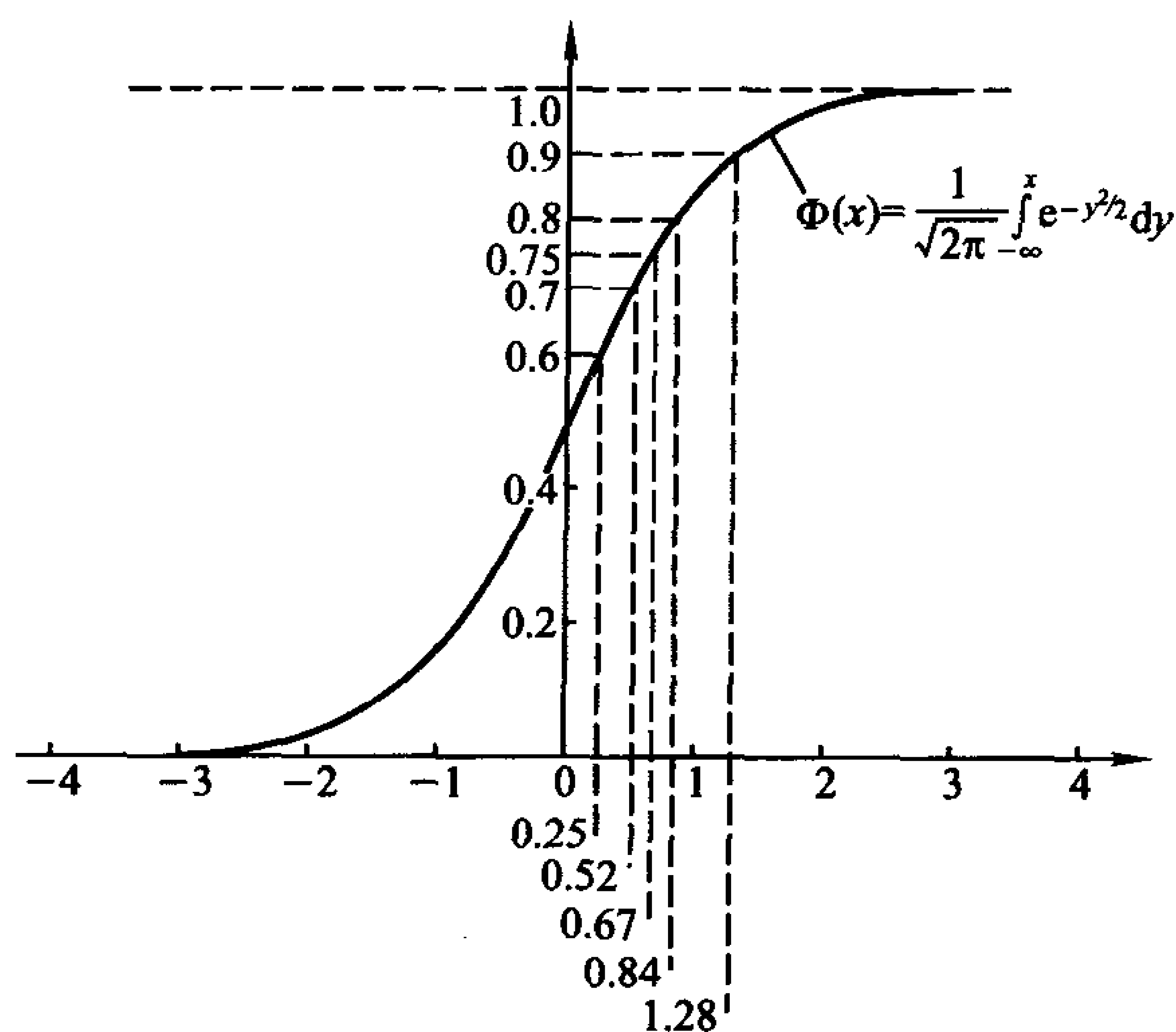


图 12 正态分布函数 $\Phi(x)$ 图

函数 $\varphi(x)$ 的图形是关于纵轴对称的钟形曲线，随着 $|x|$ 的增长下降得非常快: $\varphi(1) = 0.241\ 97, \varphi(2) = 0.053\ 991, \varphi(3) = 0.004\ 432, \varphi(4) = 0.000\ 134, \varphi(5) = 0.000\ 016$. 该曲线在 $x = 0$ 达到最大值 $(2\pi)^{-1/2} \approx 0.399$.

随着 x 的增长，函数

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t$$

曲线趋向 1 极快: $\Phi(1) = 0.841\ 345, \Phi(2) = 0.977\ 250, \Phi(3) = 0.998\ 650, \Phi(4) =$

0.999 968, $\Phi(5) = 0.999\ 997$.

关于函数 $\varphi(x)$ 和 $\Phi(x)$, 以及概率论和数理统计中其他一些基本函数的数值表, 参见 [6].

需要指出, 在进行计算时, 除函数 $\Phi(x)$ 外, 还常使用与之相近的*误差函数*

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2} \mathrm{d}t, \quad x > 0.$$

显然, 对于 $x > 0$, 有

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \ \operatorname{erf}(x) = 2\Phi(\sqrt{2}x) - 1.$$

7. 成功频率对概率的偏差满足一定要求的试验次数 在 §5 第 3 小节的最后曾经指出, 由切比雪夫不等式给出的事件

$$\left\{\omega : \left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\}$$

概率的估计是相当粗略的. 这一估计对于非负随机变量 X, 是由切比雪夫不等式

$$\mathbf{P}\{X \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}X^2}{\varepsilon^2}$$

得到的. 不过可以利用切比雪夫不等式的另一种形式

$$\mathbf{P}\{X \geqslant \varepsilon\} = \mathbf{P}\{X^{2k} \geqslant \varepsilon^{2k}\} \leqslant \frac{\mathbf{E}X^{2k}}{\varepsilon^{2k}}. \tag{33}$$

然而, 还可以更进一步, 利用切比雪夫不等式的 “指数” 形式: 对于 $X \geqslant 0$ 和 $\lambda > 0$,

$$\mathbf{P}\{X \geqslant \varepsilon\} = \mathbf{P}\{\mathrm{e}^{\lambda X} \geqslant \mathrm{e}^{\lambda\varepsilon}\} \leqslant \mathbf{E}\mathrm{e}^{\lambda(X-\varepsilon)}. \tag{34}$$

由于 $\lambda > 0$ 的任意性, 可见

$$\mathbf{P}\{X \geqslant \varepsilon\} \leqslant \inf_{\lambda>0} \mathbf{E}\mathrm{e}^{\lambda(X-\varepsilon)}. \tag{35}$$

我们讨论, 当 $X = S_n/n, S_n = \xi_1 + \cdots + \xi_n, \mathbf{P}\{\xi_i = 1\} = p, \mathbf{P}\{\xi_i = 0\} = q, i \geqslant 1$ 时, 沿此路径将导致何种结果.

记 $\varphi(\lambda) = \mathrm{e}^{\lambda\xi_1}$, 则

$$\varphi(\lambda) = 1 - p + p\mathrm{e}^{\lambda},$$

且在假设 $\xi_1, \cdots, \xi_n$ 独立的条件下, 有

$$\mathbf{E}\mathrm{e}^{\lambda S_n} = [\varphi(\lambda)]^n.$$

因此, 对于 $0 < a < 1$,

$$\mathbf{P}\left\{\frac{S_n}{n} \geqslant a\right\} \leqslant \inf_{\lambda>0} \mathbf{E}e^{\lambda\left(\frac{S_n}{n}-a\right)} = \inf_{\lambda>0} e^{-n\left[\frac{\lambda}{n}a-\ln\varphi\left(\frac{\lambda}{n}\right)\right]}$$
$$= \inf_{s>0} e^{-n[as-\ln\varphi(s)]} = e^{-n\sup\limits_{s>0}[as-\ln\varphi(s)]}. \qquad (36)$$

类似地

$$\mathbf{P}\left\{\frac{S_n}{n} \leqslant a\right\} \leqslant e^{-n\sup\limits_{s<0}[as-\ln\varphi(s)]}. \qquad (37)$$

当 $p \leqslant a \leqslant 1$ 时, 函数 $f(s) = as - \ln(1-p+pe^s)$ 在点 $s_0[f'(s_0)=0]$ 达到最大值, 其中点 s_0 决定于

$$e^{s_0} = \frac{a(1-p)}{p(1-a)}.$$

因此

$$\sup_{s>0} f(s) = H(a),$$

其中

$$H(a) = a\ln\frac{a}{p} + (1-a)\ln\frac{1-a}{1-p},$$

是前面证明局部定理时引进的函数 (第 1 小节).

这样, 当 $p \leqslant a \leqslant 1$ 时

$$\mathbf{P}\left\{\frac{S_n}{n} \geqslant a\right\} \leqslant e^{-nH(a)}, \qquad (38)$$

而由于 $H(p+x) \geqslant 2x^2, 0 \leqslant p+x \leqslant 1$, 则对于 $\varepsilon > 0, 0 \leqslant p \leqslant 1$, 有

$$\mathbf{P}\left\{\frac{S_n}{n} - p \geqslant \varepsilon\right\} \leqslant e^{-2n\varepsilon^2}. \qquad (39)$$

类似可得, 当 $a \leqslant p \leqslant 1$ 时

$$\mathbf{P}\left\{\frac{S_n}{n} \leqslant a\right\} \leqslant e^{-nH(a)}, \qquad (40)$$

从而, 对任何 $\varepsilon > 0, 0 \leqslant p \leqslant 1$, 有

$$\mathbf{P}\left\{\frac{S_n}{n} - p \leqslant -\varepsilon\right\} \leqslant e^{-2n\varepsilon^2}. \qquad (41)$$

于是

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} \leqslant 2e^{-2n\varepsilon^2}. \qquad (42)$$

由此可见, 对于任意 $0 \leqslant p \leqslant 1$, 保证不等式

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right\} \geqslant 1 - \alpha \qquad (43)$$

成立的观测次数 $n_3(\alpha)$ 决定于下面的公式:

$$n_3(\alpha) = \left[\frac{\ln(2/\alpha)}{2\varepsilon^2}\right], \tag{44}$$

其中 $[x]$ 是 x 的整数部分. 不取 "整数部分", 直接将 $n_3(\alpha)$ 与 $n_1(\alpha) = [(4\alpha\varepsilon^2)^{-1}]$ 比较, 可见

$$\frac{n_1(\alpha)}{n_3(\alpha)} = \frac{1}{4\alpha\varepsilon^2}\Big/\frac{\ln(2/\alpha)}{2\varepsilon^2} = \frac{1}{2\alpha\ln\dfrac{2}{\alpha}} \uparrow \infty, \quad \alpha \downarrow 0.$$

由此可见, 当 $\alpha \downarrow 0$ 时, 由指数型切比雪夫不等式 (34) 估计的最小必须要的观测次数, 比用一般切比雪夫不等式估计的次数更为准确, 特别是对于较小的 α. 利用将证明的关系式

$$\frac{1}{\sqrt{2\pi}}\int_x^\infty \mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y \sim \frac{1}{\sqrt{2\pi}x}\mathrm{e}^{-\frac{x^2}{2}}, \quad x \to \infty,$$

可以证明, 当 $\alpha \downarrow 0$ 时 $k^2(\alpha) \sim 2\ln(2/\alpha)$. 于是

$$\frac{n_2(\alpha)}{n_3(\alpha)} \to 1, \alpha \downarrow 0.$$

像 (38) ~ (42) 式类型的关系式, 在概率论中称做*大偏差概率的不等式*. 下面是对这一名称的解释.

利用棣莫弗 – 拉普拉斯定理, 可以简单地估计事件 $\{|S_n - np| \leqslant x\sqrt{n}\}$ 的概率, 此事件表示 S_n 对 np (数量级 $\sqrt{n}$) 的 "标准" 离差. 而不等式 (39),(41) 和 (42) 对

$$\{|S_n - np| \leqslant x\sqrt{n}\}$$

给出的估计描绘量级大于 $\sqrt{n}$ 的离差, 其数量级为 n.

我们将在第四章 §5 中, 在更一般的情形下研究关于大偏差概率的问题.

8. 练习题

1. 设 $n = 100, p = 1/10, p = 2/10, p = 3/10, p = 4/10, p = 5/10$. 利用 (例如文献 [6] 中的) 二项分布以及泊松分布的数值表①, 将概率

$$\mathbf{P}\{10 < S_{100} \leqslant 12\}, \quad \mathbf{P}\{20 < S_{100} \leqslant 22\},$$
$$\mathbf{P}\{33 < S_{100} \leqslant 35\}, \quad \mathbf{P}\{40 < S_{100} \leqslant 42\},$$
$$\mathbf{P}\{50 < S_{100} \leqslant 52\}$$

与正态逼近和泊松逼近的相应数值进行比较.

2. 设 $p = 1/2$, 而 $Z_n = 2S_n - n$ (n 组试验中 1 比 0 多出的个数). 证明

$$\sup_j |\sqrt{\pi n}\mathbf{P}\{Z_{2n} = j\} - \mathrm{e}^{-j^2/4^n}| \to 0, \quad n \to \infty.$$

①亦可利用: 中国科学院数学研究所编的《常用数理统计表》, 科学出版社, 1974 年. —— 译者

3. 证明泊松定理 (对于 $p=\lambda/n$) 的收敛速度为

$$\sup_k \left|P_n(k)-\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}\right| \leqslant \frac{\lambda^2}{n}.$$

(证明可以参见第三章 §12.)

§7. 伯努利概型中“成功”概率的估计

1. “成功”概率估计的概念和性质 (相合性, 无偏性和有效性). 以上讨论的伯努利概型

$$(\Omega,\mathscr{A},\mathbf{P}),\Omega=\{\omega:\omega=(x_1,\cdots,x_n),x_i=0,1\},$$
$$\mathscr{A}=\{A:A\subseteq\Omega\},\mathbf{P}(\{\omega\})=p(\omega),$$
$$p(\omega)=p^{\sum x_i}(1-p)^{n-\sum x_i},$$

假设 p (“成功” 的概率) 的数值已知.

现在假设 p 事先未知, 并希望根据对试验结局的观测结果, 或由对随机变量 $\xi_1,\cdots,\xi_n$ 的观测结果来确定 p, 其中 $\xi_i(\omega)=x_i$. 这是数理统计的典型问题之一, 有不同的提法. 我们下面讨论问题的两种提法:*估计问题*和*建立置信区间问题*.

沿用数理统计中普遍采用的记号, 未知参数 p 记作 θ, 并认为是验前的 (a priori), 且 θ 的值属于集合 $\Theta=[0,1]$. 通常称 “由

$$\mathscr{E}=(\Omega,\mathscr{A},\mathbf{P}_\theta,\theta\in\Theta),\quad \mathbf{P}_\theta(\{\omega\})=\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

确定一个 (对应于 ‘n 次独立试验’ 且 ‘成功’ 概率为 $\theta\in\Theta$ 的) *概率 – 统计模型*”, 而任何在 Θ 中取值的函数 $T_n=T_n(\omega)$ 称做*估计量*.

如果设

$$S_n=\xi_1+\cdots+\xi_n,\quad T_n^*=\frac{S_n}{n},$$

则由大数定律可见, 估计量 T_n^* 称做*相合的*, 若对于任意 $\varepsilon>0$, 有

$$\mathbf{P}_\theta\{|T_n^*-\theta|\geqslant\varepsilon\}\to 0,\quad n\to\infty. \tag{1}$$

此外, 估计量 T_n^* 称做*无偏的*, 如果对于任意 $\theta\in\Theta$,

$$\mathbf{E}_\theta T_n^*=\theta, \tag{2}$$

其中 $\mathbf{E}_\theta$ 是对应于概率 $\mathbf{P}_\theta$ 的数学期望.

估计量的无偏性是一条很自然的性质, 它反映如下事实: 由任何*合理的*估计量, 至少在平均意义下都应当得到所期待的结果. 不过, 估计量 T_n^* 并非唯一无偏估计量. 例如, 对于 $b_1+\cdots+b_n=n$, 任何一个估计量

$$T_n=\frac{b_1\xi_1+\cdots+b_n\xi_n}{n}$$

都是无偏的. 这些估计量都服从大数定律 (1) (至少对于 $|b_i| \leqslant K < \infty$), 从而, 这些估计量 T_n 也和 T_n^* 一样是 "好" 估计量.

于是, 产生一个问题: 如何比较不同的无偏估计量, 它们之中哪一个应称做最好的、最优的.

按估计量的本身的含义, 自然应当认为估计量对被估计的参数的偏差越小越好. 基于这样的考虑, 称估计量 $\widetilde{T}_n$ (在无偏估计 T_n 类中) 为**有效的**, 如果

$$\mathbf{D}_\theta \widetilde{T}_n = \inf_{T_n} \mathbf{D}_\theta T_n, \quad \theta \in \Theta, \tag{3}$$

其中 $\mathbf{D}_\theta T_n$ 是估计量 T_n 的方差, 即 $\mathbf{E}_\theta (T_n - \theta)^2$.

现在证明上面所考虑的估计量 T_n^* 是有效估计量. 事实上, 有

$$\mathbf{D}_\theta T_n^* = \mathbf{D}_\theta \left(\frac{S_n}{n} \right) = \frac{\mathbf{D}_\theta S_n}{n^2} = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}. \tag{4}$$

因此, 为证明估计量 T_n^* 有效, 只需证明

$$\inf_{T_n} \mathbf{D}_\theta T_n \geqslant \frac{\theta(1-\theta)}{n}. \tag{5}$$

对于 $\theta = 0$ 和 1, 不等式显然. 现在设 $\theta \in (0,1)$ 且

$$p_\theta(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}.$$

显然 $\mathbf{P}_\theta(\{\omega\}) = p_\theta(\omega)$, 其中

$$p_\theta(\omega) = \prod_{i=1}^n p_\theta(x_i).$$

记

$$L_\theta(\omega) = \ln p_\theta(\omega).$$

那么

$$L_\theta(\omega) = \ln\theta \sum_{i=1}^n x_i + \ln(1-\theta) \sum_{i=1}^n (1 - x_i),$$

$$\frac{\partial L_\theta(\omega)}{\partial \theta} = \frac{\sum\limits_{i=1}^n (x_i - \theta)}{\theta(1-\theta)}.$$

因为

$$1 \equiv \mathbf{E}_\theta 1 = \sum_\omega p_\theta(\omega)$$

且由估计量 T_n 的无偏性

$$\theta \equiv \mathbf{E}_\theta T_n = \sum_\omega T_\theta(\omega) p_\theta(\omega),$$

所以经对 θ 求导, 得

$$0=\sum_{\omega}\frac{\partial p_\theta(\omega)}{\partial\theta}=\sum_{\omega}\frac{\frac{\partial p_\theta(\omega)}{\partial\theta}}{p_\theta(\omega)}p_\theta(\omega)=\mathbf{E}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right],$$

$$1=\sum_{\omega}T_n\frac{\frac{\partial p_\theta(\omega)}{\partial\theta}}{p_\theta(\omega)}p_\theta(\omega)=\mathbf{E}_\theta\left[T_n\frac{\partial L_\theta(\omega)}{\partial\theta}\right].$$

因此,

$$1=\mathbf{E}_\theta\left[(T_n-\theta)\frac{\partial L_\theta(\omega)}{\partial\theta}\right],$$

而根据柯西 – 布尼亚科夫斯基不等式

$$1\leqslant\mathbf{E}_\theta(T_n-\theta)^2\times\mathbf{E}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right]^2.$$

于是,

$$\mathbf{E}_\theta(T_n-\theta)^2\geqslant\frac{1}{I_n(\theta)},\tag{6}$$

其中

$$I_n(\theta)=\mathbf{E}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right]^2$$

称做费希尔信息量.

由 (6) 式得无偏估计量的所谓拉奥 – 克拉默 (C. R. Rao – G. Cramér) 不等式的特殊情形:

$$\inf_{T_n}\mathbf{D}_\theta T_n\geqslant\frac{1}{I_n(\theta)}.\tag{7}$$

对于所讨论的情形, 有

$$I_n(\theta)=\mathbf{E}_\theta\left[\frac{\partial L_\theta(\omega)}{\partial\theta}\right]^2=\mathbf{E}_\theta\left[\frac{\sum\limits_{i=1}^{n}(\xi_i-\theta)}{\theta(1-\theta)}\right]^2=\frac{n\theta(1-\theta)}{[\theta(1-\theta)]^2}=\frac{n}{\theta(1-\theta)},$$

从而不等式 (5) 得证. 如我们曾提到的那样, 由此可见 $T_n^*=S_n/n$ 是未知参数 θ 的有效估计量.

2. "成功" 概率的置信区间 显然, 把 T_n^* 当作 θ 的 "点" 估计量, 我们就犯了某种错误. 甚至有可能出现这样的情形, 由观测数据 $x_1,\cdots,x_n$ 计算的数值 T_n^*, 对 θ 的真值有相当大的偏差. 因此最好再指出误差的大小.

不能指望对所有基本事件 ω, $T_n^*=T_n^*(\omega)$ 都能与未知参数 θ 的真值差异甚小, 这也是毫无意义的. 不过, 由大数定律, 知对于充分大的 n 和任意 $\delta>0$, 事件 $\{|\theta-T_n^*|>\delta\}$ 的概率都充分小.

根据切比雪夫不等式, 有

$$\mathbf{P}_\theta\{|\theta - T_n^*| > \delta\} \leqslant \frac{\mathbf{D}_\theta T_n^*}{\delta^2} = \frac{\theta(1-\theta)}{n\delta^2},$$

因此, 对于任意 $\lambda > 0$, 有

$$\mathbf{P}_\theta\left\{|\theta - T_n^*| \leqslant \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\right\} \geqslant 1 - \frac{1}{\lambda^2}.$$

例如, 取 $\lambda = 3$, 则事件

$$\left\{|\theta - T_n^*| \leqslant 3\sqrt{\frac{\theta(1-\theta)}{n}}\right\}$$

出现的概率 $\mathbf{P}_\theta$ 大于 0.8888 ($1 - 1/3^2 = 8/9 \approx 0.8888$). 特别, 因为 $\{\theta(1-\theta) \leqslant 1/4\}$, 故事件

$$\left\{|\theta - T_n^*| \leqslant \frac{3}{2\sqrt{n}}\right\}$$

出现的概率 $\mathbf{P}_\theta$ 大于 0.8888.

于是,

$$\mathbf{P}_\theta\left\{|\theta - T_n^*| \leqslant \frac{3}{2\sqrt{n}}\right\} = \mathbf{P}_\theta\left\{T_n^* - \frac{3}{2\sqrt{n}} \leqslant \theta \leqslant T_n^* + \frac{3}{2\sqrt{n}}\right\} \geqslant 0.8888.$$

换句话说, 可以断定 “未知参数 θ 的真值” 以大于 0.8888 的概率属于区间

$$\left[T_n^* - \frac{3}{2\sqrt{n}}, T_n^* + \frac{3}{2\sqrt{n}}\right].$$

有时将此命题简单地表示为

$$\theta \simeq T_n^* \pm \frac{3}{2\sqrt{n}} (\geqslant 88\%),$$

其中 “$\geqslant 88\%$” 表示 “在概率不小于 88% 的情形下”.

区间

$$\left[T_n^* - \frac{3}{2\sqrt{n}}, T_n^* + \frac{3}{2\sqrt{n}}\right]$$

就是 “未知参数的置信区间” 的一个例子.

定义 称形如

$$[\psi_1(\omega), \psi_2(\omega)]$$

的集合为*置信度 $1-\delta$ 的置信区间* (或显著性水平为 δ 的置信区间), 如果对于一切 $\theta \in \Theta$, 有

$$\mathbf{P}_\theta\{\psi_1(\theta) \leqslant \theta \leqslant \psi_2(\theta)\} \geqslant 1 - \delta,$$

其中 $\psi_1(\omega)$ 和 $\psi_2(\omega)$ 是基本事件的两个函数.

上面的讨论表明, 区间

$$\left[T_n^* - \frac{\lambda}{2\sqrt{n}}, T_n^* + \frac{\lambda}{2\sqrt{n}}\right]$$

的置信度为 $1-1/\lambda^2$. 实际上, 置信度要远远高些, 因为由切比雪夫不等式只能得到事件概率的粗略估计. 为得到更精确的结果, 注意到

$$\left\{\omega : |\theta - T_n^*| \leqslant \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\right\} = \{\omega : \psi_1(T_n^*, n) \leqslant \theta \leqslant \psi_2(T_n^*, n)\},$$

其中 $\psi_1 = \psi_1(T_n^*, n)$ 和 $\psi_2 = \psi_2(T_n^*, n)$ 是如下二次方程

$$(\theta - T_n^*)^2 = \frac{\lambda^2}{n}\theta(1-\theta)$$

的根, 该方程描绘图 13 所示的椭圆. 现在记

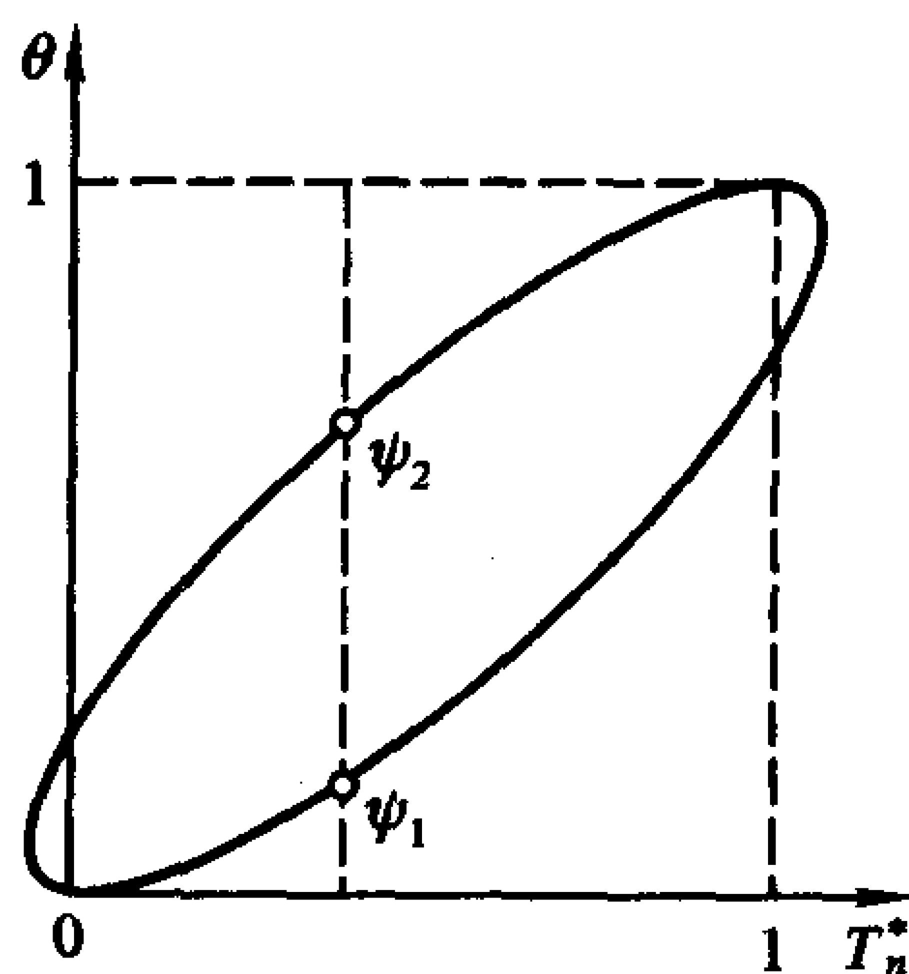


图 13

$$F_\theta^n(x) = \mathbf{P}_\theta\left\{\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leqslant x\right\}.$$

那么, 由 §6(24) 式, 可见

$$\sup_x |F_\theta^n(x) - \Phi(x)| \leqslant \frac{1}{\sqrt{n\theta(1-\theta)}}.$$

因此, 假如事先已知

$$0 < \Delta \leqslant \theta \leqslant 1 - \Delta < 1,$$

其中 Δ 是某一常数, 则

$$\sup_x |F_\theta^n(x) - \Phi(x)| \leqslant \frac{1}{\Delta\sqrt{n}}.$$

从而

$$\begin{aligned}\mathbf{P}_\theta\{\psi_1(T_n^*, n) \leqslant \theta \leqslant \psi_2(T_n^*, n)\} &= \mathbf{P}_\theta\left\{|\theta - T_n^*| \leqslant \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\right\} \\ &= \mathbf{P}_\theta\left\{\frac{|S_n - n\theta|}{\sqrt{n\theta(1-\theta)}} \leqslant \lambda\right\} \geqslant [2\Phi(\lambda) - 1] - \frac{2}{\Delta\sqrt{n}}.\end{aligned}$$

设 λ^* 是满足

$$[2\Phi(\lambda) - 1] - \frac{2}{\Delta\sqrt{n}} \leqslant 1 - \delta^*$$

的最小 λ 值, 其中 δ^* 是给定的显著性水平. 记 $\delta = \delta^* - 2/(\Delta\sqrt{n})$, 则 λ^* 是如下方程的根:

$$\Phi(\lambda) = 1 - \frac{\delta}{2}.$$

当 n 较大时, 可以忽略项 $2/(\Delta\sqrt{n})$, 认为 λ^* 满足关系式

$$\Phi(\lambda^*) = 1 - \frac{\delta^*}{2}.$$

例如, 若 $\lambda^* = 3$, 则 $1 - \delta_*^* = 0.9973\cdots$ 因此大致以概率 0.9973, 有

$$T_n^* - 3\sqrt{\frac{\theta(1-\theta)}{n}} \leqslant \theta \leqslant T_n^* + 3\sqrt{\frac{\theta(1-\theta)}{n}}, \tag{8}$$

而经迭代和忽略量级为 $O(n^{-3/4})$ 的项, 得

$$T_n^* - 3\sqrt{\frac{T_n^*(1-T_n^*)}{n}} \leqslant \theta \leqslant T_n^* + 3\sqrt{\frac{T_n^*(1-T_n^*)}{n}}. \tag{9}$$

由此可见, 对于充分大 n 的, 置信区间

$$\left[T_n^* - \frac{3}{2\sqrt{n}}, T_n^* + \frac{3}{2\sqrt{n}}\right]$$

的置信度为 0.9973, 然而由切比雪夫不等式得到的置信度只有 0.8888.

由此可以得到如下实用的结果. 假如进行大数量 N 系列试验, 每系列试验根据 n 次观测的结果估计参数 θ. 那么, 平均在 97.33% 的情形下, 每一系列 n 次试验, 估计量与参数真值的差别不大于 $3/(2\sqrt{n})$. (关于这一结果, 亦可参见 §5 的末尾.)

3. 练习题

1. 假设事先已知参数 θ 在 $\Theta_0 \subseteq [0,1]$ 中取值. 说明何时对于只在 Θ_0 中取值的参数 θ 存在无偏估计.

2. 在上题的条件下求拉奥 – 克拉默不等式, 并讨论估计量的有效性.

3. 在第 1 题的条件下, 讨论建立 θ 的置信区间的问题.

4. 在 §2 第 5 题的条件下, 假设 $N \gg M, N \gg n$, 讨论估计量 $\widehat{N}$ 的无偏性和有效性. 建立 θ 的置信区间 (见 (8) 式和 (9) 式), N 的置信区间 $[\widehat{N} - a(\widehat{N}), \widehat{N} + b(\widehat{N})]$, 使

$$\mathbf{P}_{N,M;n}\{\widehat{N} - a(\widehat{N}) \leqslant N \leqslant \widehat{N} + b(\widehat{N})\} = 1 - \alpha,$$

其中 α 是某一较小的数.

§8. 关于分割的条件概率与条件数学期望

1. 条件概率 设 $(\Omega, \mathscr{A}, \mathbf{P})$ 是概率空间, 而

$$\mathscr{D} = \{D_1, \cdots, D_k\}$$

是 Ω 某一分割: $D_i \in \mathscr{A}, \mathbf{P}(D_i) > 0, i = 1, \cdots, k; D_1 + \cdots + D_k = \Omega$. 其次, 设事件 $A \in \mathscr{A}$, 而 $\mathbf{P}(A|D_i) > 0$, 是事件 A 关于事件 D_i 的条件概率.

一组条件概率 $\{\mathbf{P}(A|D_i), i=1,\cdots,k\}$ 可以与随机变量

$$\pi(\omega)=\sum_{i=1}^{k}\mathbf{P}(A|D_i)I_{D_i}(\omega) \tag{1}$$

相联系 (与 §4 (5) 式比较), 且 $\pi(\omega)$ 在原子 D_i 上取 $\mathbf{P}(A|D_i)$ 为值. 为强调 $\pi(\omega)$ 确实与此分割 $\mathscr{D}$ 相联系, 将其记作

$$\mathbf{P}(A|\mathscr{D}) \quad 或 \quad \mathbf{P}(A|\mathscr{D})(\omega),$$

并称之为**事件 A 关于分割 $\mathscr{D}$ 的条件概率**.

这一概念, 以及将要引进的关于 $\sigma-$ 代数的条件概率的概念, 在概率论中起着重要的作用, 下面将逐步展开叙述.

条件概率的如下两条明确的性质:

$$\mathbf{P}(A+B|\mathscr{D})=\mathbf{P}(A|\mathscr{D})+\mathbf{P}(B|\mathscr{D}); \tag{2}$$

如果 $\mathscr{D}$ 是只含 Ω 中一个集合的平凡分割, 则

$$\mathbf{P}(A|\mathscr{D})=\mathbf{P}(A). \tag{3}$$

把条件概率 $\mathbf{P}(A|\mathscr{D})$ 定义为**随机变量**, 就可以考虑其数学期望, 利用数学期望可以用如下紧凑的形式将 §3 中**全概率公式** (3) 写成:

$$\mathbf{E}\mathbf{P}(A|\mathscr{D})=\mathbf{P}(A). \tag{4}$$

事实上, 由于

$$\mathbf{P}(A|\mathscr{D})(\omega)=\sum_{i=1}^{k}\mathbf{P}(A|D_i)I_{D_i}(\omega),$$

则根据数学期望的定义 (见 §4 (5) 和 (6) 式), 有

$$\mathbf{E}\mathbf{P}(A|\mathscr{D})=\sum_{i=1}^{k}\mathbf{P}(A|D_i)\mathbf{P}(D_i)=\sum_{i=1}^{k}\mathbf{P}(AD_i)=\mathbf{P}(A).$$

现在, 设 $\eta=\eta(\omega)$ 是以大于 0 的概率取 $y_1,\cdots,y_k$ 为值的随机变量:

$$\eta(\omega)=\sum_{j=1}^{k}y_jI_{D_j}(\omega),$$

其中 $D_j=\{\omega:\eta(\omega)=y_j\}$. 分割 $\mathscr{D}_\eta=\{D_1,\cdots,D_k\}$ 称做随机变量 η 诱导的分割. 条件概率 $\mathbf{P}(A|\mathscr{D}_\eta)$ 或 $\mathbf{P}(A|\mathscr{D}_\eta)(\omega)$ 称做**事件 A 关于随机变量 η 的条件概率**. 我们把 $\mathbf{P}(A|\eta=y_j)$ 也理解为条件概率 $\mathbf{P}(A|D_j)$, 其中 $D_j=\{\omega:\eta(\omega)=y_j\}$.

类似地, 若 $\eta_1,\eta_2,\cdots,\eta_m$ 是随机变量, 而 $\mathscr{D}_{\eta_1,\eta_2,\cdots,\eta_m}$ 是诱导的以

$$D_{y_1,y_2,\cdots,y_m}=\{\omega:\eta_1=y_1,\eta_2=y_2,\cdots,\eta_m=y_m\}$$

为原子的分割, 则 $\mathbf{P}(A|\mathscr{D}_{\eta_1,\eta_2,\cdots,\eta_m})$ 记作 $\mathbf{P}(A|\eta_1,\eta_2,\cdots,\eta_m)$, 并称为事件 A 关于随机变量 $\eta_1,\eta_2,\cdots,\eta_m$ 的条件概率.

例 1　设 ξ 和 η 是两个独立同分布随机变量, 分别以概率 p 和 q 取 1 和 0 为值. 对于 $k=0,1,2$, 我们现在求事件 $A=\{\omega:\xi+\eta=k\}(k=0,1,2)$ 关于 η 的条件概率 $\mathbf{P}(A|\eta)$.

为此, 首先注意到有用的一般事实: 如果 ξ 和 η 是两个独立随机变量, 则当 $\mathbf{P}\{\eta=y\}>0$ 时, 有

$$\mathbf{P}\{\xi+\eta=z|\eta=y\}=\mathbf{P}\{\xi+y=z\}. \tag{5}$$

事实上,

$$\begin{aligned}\mathbf{P}\{\xi+\eta=z|\eta=y\}&=\frac{\mathbf{P}\{\xi+\eta=z,\eta=y\}}{\mathbf{P}\{\eta=y\}}=\frac{\mathbf{P}\{\xi+y=z,\eta=y\}}{\mathbf{P}\{\eta=y\}}\\&=\frac{\mathbf{P}\{\xi+y=z\}\mathbf{P}\{\eta=y\}}{\mathbf{P}\{\eta=y\}}=\mathbf{P}\{\xi+y=z\}.\end{aligned}$$

由此公式, 可见

$$\begin{aligned}\mathbf{P}(A|\eta)(\omega)&=\mathbf{P}(\xi+\eta=k|\eta)(\omega)\\&=\mathbf{P}(\xi+\eta=k|\eta=0)I_{\{\eta=0\}}(\omega)+\mathbf{P}(\xi+\eta=k|\eta=1)I_{\{\eta=1\}}(\omega)\\&=\mathbf{P}(\xi=k)I_{\{\eta=0\}}(\omega)+\mathbf{P}(\xi=k-1)I_{\{\eta=1\}}(\omega).\end{aligned}$$

于是

$$\mathbf{P}(\xi+\eta=k|\eta)(\omega)=\begin{cases}qI_{\{\eta=0\}}(\omega), & 若\ k=0,\\ pI_{\{\eta=0\}}(\omega)+qI_{\{\eta=1\}}(\omega), & 若\ k=1,\\ pI_{\{\eta=1\}}(\omega), & 若\ k=2,\end{cases} \tag{6}$$

或同样地

$$\mathbf{P}(\xi+\eta=k|\eta)=\begin{cases}q(1-\eta), & 若\ k=0,\\ p(1-\eta)+q\eta, & 若\ k=1,\\ p\eta, & 若\ k=2.\end{cases} \tag{7}$$

2. 条件数学期望　设 $\xi=\xi(\omega)$ 是随机变量, 其值域为 $X=\{x_1,\cdots,x_l\}$:

$$\xi=\sum_{j=1}^{l}x_jI_{A_j},\quad A_j=\{\omega:\xi=x_j\},$$

而 $\mathscr{D} = \{D_1, \cdots, D_k\}$ 是某个分割. 曾经定义了 ξ 关于概率 $\mathbf{P}(A_j)(j = 1, \cdots, l)$ 的数学期望:

$$\mathbf{E}\xi = \sum_{j=1}^{l} x_j \mathbf{P}(A_j). \tag{8}$$

自然类似地利用概率 $\mathbf{P}(A_j|\mathscr{D})$, 定义随机变量 ξ 关于分割 $\mathscr{D}$ 的条件数学期望, 记作 $\mathbf{E}(\xi|\mathscr{D})$ 或 $\mathbf{E}(\xi|\mathscr{D})(\omega)$:

$$\mathbf{E}(\xi|\mathscr{D}) = \sum_{j=1}^{l} x_j \mathbf{P}(A_j|\mathscr{D}). \tag{9}$$

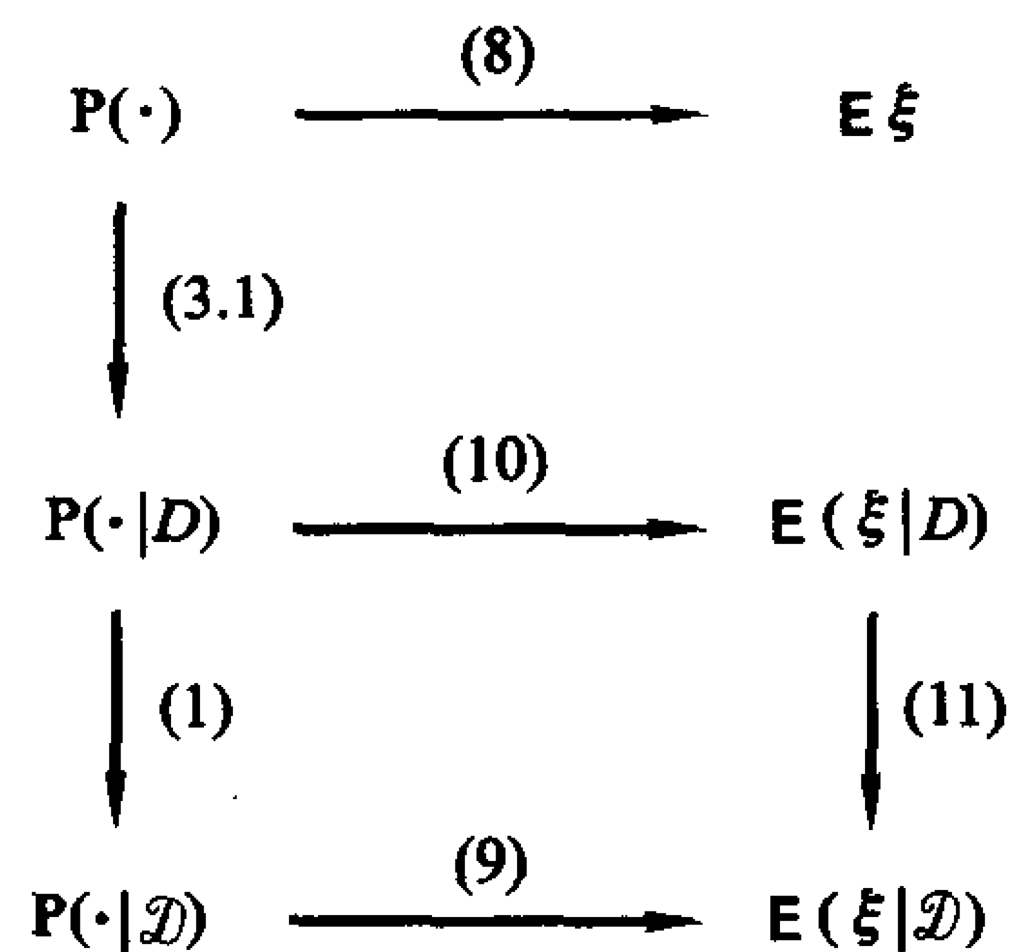


图 14

根据这一定义, 条件数学期望 $\mathbf{E}(\xi|\mathscr{D})(\omega)$ 是一随机变量, 对于属于同一原子 D_i 的基本事件 ω, 取同一个值

$$\sum_{j=1}^{l} x_j \mathbf{P}(A_j|D_i).$$

这一事实说明, 可以由另一途径定义条件数学期望. 具体地说, 首先由公式

$$\mathbf{E}(\xi|D_i) = \sum_{j=1}^{l} x_j \mathbf{P}(A_j|D_i) \left(= \frac{\mathbf{E}(\xi I_{D_i})}{\mathbf{P}(D_i)}\right) \tag{10}$$

定义随机变量 ξ 关于事件 D_i 的条件数学期望, 然后按定义设

$$\mathbf{E}(\xi|\mathscr{D})(\omega) = \sum_{i=1}^{l} \mathbf{E}(\xi|D_i) I_{D_i}(\omega) \tag{11}$$

(参见流程图 14).

注意, $\mathbf{E}(\xi|D)$ 和 $\mathbf{E}(\xi|\mathscr{D})(\omega)$ 的值与随机变量 ξ 的表示方法无关.

下面指出的条件数学期望的性质, 可以直接由其定义得到:

$$\mathbf{E}(a\xi + b\eta|\mathscr{D}) = a\mathbf{E}(\xi|\mathscr{D}) + b\mathbf{E}(\eta|\mathscr{D})(a, b \text{ 是常数}); \tag{12}$$

$$\mathbf{E}(\xi|\Omega) = \mathbf{E}\xi; \tag{13}$$

$$\mathbf{E}(C|\mathscr{D}) = C, \quad C \text{ 是常数}. \tag{14}$$

如果 $\xi = I_A(\omega)$, 则

$$\mathbf{E}(\xi|\mathscr{D}) = \mathbf{P}(A|\mathscr{D}). \tag{15}$$

其中, 最后一条性质表明, 由条件数学期望的性质, 可以直接得出条件概率的性质.

另一重要的性质是全概率公式 (4) 的推广:

$$\mathbf{E}\mathbf{E}(\xi|\mathscr{D}) = \mathbf{E}\xi. \tag{16}$$

为证明此式, 只需注意到, 由 (4) 式, 可见.

$$\mathbf{E}\mathbf{E}(\xi|\mathscr{D}) = \mathbf{E}\sum_{j=1}^{l} x_j \mathbf{P}(A_j|\mathscr{D}) = \mathbf{E}\sum_{j=1}^{l} x_j \mathbf{E}\mathbf{P}(A_j|\mathscr{D}) = \sum_{j=1}^{l} x_j \mathbf{P}(A_j) = \mathbf{E}\xi.$$

设 $\mathscr{D} = \{D_1, \cdots, D_k\}$ 是某个分割, 而 $\eta = \eta(\omega)$ 是某一随机变量. 我们称随机变量 η 关于此分割为可测的或 $\mathscr{D}-$ 可测的, 如果 $\mathscr{D}_\eta \preccurlyeq \mathscr{D}$, 即 $\eta = \eta(\omega)$ 可以表示为

$$\eta(\omega) = \sum_{i=1}^{k} y_i I_{D_i}(\omega),$$

其中有些 y_i 值可能相等. 换句话说, 随机变量 η 为 $\mathscr{D}-$ 可测的, 当且仅当它在分割 $\mathscr{D}$ 的原子上取常数值.

例 2 如果 $\mathscr{D}$ 为平凡分割: $\mathscr{D} = \{\Omega\}$, 则随机变量 η 为 $\mathscr{D}-$ 可测的, 当且仅当 $\eta \equiv C$, 其中 C 为常数. 任何随机变量 η 关于分割 $\mathscr{D}_\eta-$ 可测.

假设随机变量 η 为 $\mathscr{D}-$ 可测, 那么

$$\mathbf{E}(\xi\eta|\mathscr{D}) = \eta\mathbf{E}(\xi|\mathscr{D}), \tag{17}$$

特别

$$\mathbf{E}(\eta|\mathscr{D}) = \eta \ (\mathbf{E}(\eta|\mathscr{D}_\eta) = \eta). \tag{18}$$

为证明 (17) 式注意到, 如果

$$\xi = \sum_{j=1}^{l} x_j I_{A_j},\ A_j = \{\omega : \xi = x_j\},$$

则

$$\xi\eta = \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i I_{A_j D_i},$$

即

$$\begin{aligned}\mathbf{E}(\xi\eta|\mathscr{D}) &= \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i \mathbf{P}(A_j D_i|\mathscr{D}) = \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i \sum_{m=1}^{k} \mathbf{P}(A_j D_i|D_m) I_{D_m}(\omega) \\ &= \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i \mathbf{P}(A_j D_i|D_i) I_{D_i}(\omega) = \sum_{j=1}^{l}\sum_{i=1}^{k} x_j y_i \mathbf{P}(A_j|D_i) I_{D_i}(\omega).\end{aligned} \tag{19}$$

另一方面, 考虑到 $I_{D_i}^2 = I_{D_i}$ 和 $I_{D_i} I_{D_m} = 0, i \neq m$, 有

$$\begin{aligned}\eta\mathbf{E}(\xi|\mathscr{D}) &= \left[\sum_{i=1}^{k} y_i I_{D_i}(\omega)\right] \times \left[\sum_{j=1}^{l} x_j \mathbf{P}(A_j|\mathscr{D})\right] \\ &= \left[\sum_{i=1}^{k} y_i I_{D_i}(\omega)\right] \times \sum_{m=1}^{k} \left[\sum_{j=1}^{l} x_j \mathbf{P}(A_j|D_m)\right] \times I_{D_m}(\omega) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{l} x_j y_i \mathbf{P}(A_j|D_i) I_{D_i}(\omega),\end{aligned}$$

于是, (19) 式与 (17) 式一同得证.

我们再证明条件数学期望一条重要性质. 设 $\mathscr{D}_1$ 和 $\mathscr{D}_2$ 是两个分割, 且 $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2$ (分割 $\mathscr{D}_1$ 比分割 $\mathscr{D}_2$ "细小"). 那么, "望远" 性质成立:

$$\mathbf{E}[\mathbf{E}(\xi|\mathscr{D}_2)\mathscr{D}_1] = \mathbf{E}(\xi|\mathscr{D}_1). \tag{20}$$

为证明 (20) 式, 记 $\mathscr{D}_1 = \{D_{11}, \cdots, D_{1m}\}, \mathscr{D}_2 = \{D_{21}, \cdots, D_{2n}\}$. 那么, 若

$$\xi = \sum_{j=1}^{l} x_j I_{A_j},$$

则

$$\mathbf{E}(\xi|\mathscr{D}_2) = \sum_{j=1}^{l} x_j \mathbf{P}(A_j|\mathscr{D}_2),$$

且只需证明

$$\mathbf{E}[\mathbf{P}(A_j|\mathscr{D}_2)|\mathscr{D}_1] = \mathbf{P}(A_j|\mathscr{D}_1). \tag{21}$$

由于

$$\mathbf{P}(A_j|\mathscr{D}_2) = \sum_{q=1}^{n} \mathbf{P}(A_j|D_{2q}) I_{D_{2q}},$$

则

$$\begin{aligned}\mathbf{E}[\mathbf{P}(A_j|\mathscr{D}_2)|\mathscr{D}_1] &= \sum_{q=1}^{n} \mathbf{P}(A_j|D_{2q})\mathbf{P}(D_{2q}|\mathscr{D}_1) \\ &= \sum_{q=1}^{n} \mathbf{P}(A_j|D_{2q}) \left[\sum_{p=1}^{m} \mathbf{P}(D_{2q}|D_{1p}) I_{D_{1p}}\right] \\ &= \sum_{p=1}^{m} I_{D_{1p}} \sum_{q=1}^{n} \mathbf{P}(A_j|D_{2q})\mathbf{P}(D_{2q}|D_{1p}) \\ &= \sum_{p=1}^{m} I_{D_{1p}} \sum_{\{q:D_{2q}\subseteq D_{1p}\}} \mathbf{P}(A_j|D_{2q})\mathbf{P}(D_{2q}|D_{1p}) \\ &= \sum_{p=1}^{m} I_{D_{1p}} \sum_{\{q:D_{2q}\subseteq D_{1p}\}} \frac{\mathbf{P}(A_j D_{2q})}{\mathbf{P}(D_{2q})} \times \frac{\mathbf{P}(D_{2q})}{\mathbf{P}(D_{1p})} \\ &= \sum_{p=1}^{m} I_{D_{1p}} \mathbf{P}(A_j|D_{1p}) = \mathbf{P}(A_j|\mathscr{D}_1),\end{aligned}$$

于是, (21) 式得证.

假如分割 $\mathscr{D}$ 是由随机变量 $\eta_1, \cdots, \eta_k$ 诱导的 ($\mathscr{D} = \mathscr{D}_{\eta_1,\cdots,\eta_k}$), 则条件数学期望 $\mathbf{E}(\xi|\mathscr{D}_{\eta_1,\cdots,\eta_k})$ 称做随机变量 ξ 关于 $\eta_1, \cdots, \eta_k$ 的条件数学期望, 记作 $\mathbf{E}(\xi|\eta_1, \cdots, \eta_k)$ 或 $\mathbf{E}(\xi|\eta_1, \cdots, \eta_k)(\omega)$.

直接由 $\mathbf{E}(\xi|\eta)$ 的定义可见, 如果 ξ 和 η 独立, 则

$$\mathbf{E}(\xi|\eta) = \mathbf{E}\xi. \tag{22}$$

由 (18) 式, 可见

$$\mathbf{E}(\eta|\eta) = \eta. \tag{23}$$

性质 (22) 有如下推广. 设随机变量 ξ 与分割 $\mathscr{D}$ 独立 (即对于任意 $D_i \in \mathscr{D}$, 随机变量 ξ 与 I_{D_i} 独立). 那么,

$$\mathbf{E}(\xi|\mathscr{D}) = \mathbf{E}\xi. \tag{24}$$

作为特殊情形, 由 (20) 式, 可以得到如下有用的公式:

$$\mathbf{E}[\mathbf{E}(\xi|\eta_1, \eta_2)|\eta_1] = \mathbf{E}(\xi|\eta_1). \tag{25}$$

例 3　对于例 1 中的随机变量 ξ 和 η, 求 $\mathbf{E}(\xi + \eta|\eta)$. 由 (22) 式和 (23) 式, 有

$$\mathbf{E}(\xi + \eta|\eta) = \mathbf{E}\xi + \eta = p + \eta.$$

由 (8) 式亦可得到此结果:

$$\mathbf{E}(\xi + \eta|\eta) = \sum_{k=0}^{2} k\mathbf{P}(\xi + \eta = k|\eta) = p(1 - \eta) + q\eta + 2p\eta = p + \eta.$$

例 4 设 ξ 和 η 是独立同分布随机变量, 验证,

$$\mathbf{E}(\xi|\xi+\eta)=\mathbf{E}(\eta|\xi+\eta)=\frac{\xi+\eta}{2}. \tag{26}$$

事实上, 为简便计, 假设 ξ 和 η 取 $1,2,\cdots,m$ 为值; 对于 $1\leqslant k\leqslant m, 2\leqslant l\leqslant 2m$, 有

$$\begin{aligned}\mathbf{P}(\xi=k|\xi+\eta=l)&=\frac{\mathbf{P}\{\xi=k,\xi+\eta=l\}}{\mathbf{P}\{\xi+\eta=l\}}=\frac{\mathbf{P}\{\xi=k,\eta=l-k\}}{\mathbf{P}\{\xi+\eta=l\}}\\&=\frac{\mathbf{P}\{\xi=k\}\mathbf{P}\{\eta=l-k\}}{\mathbf{P}\{\xi+\eta=l\}}=\frac{\mathbf{P}\{\eta=k\}\mathbf{P}\{\xi=l-k\}}{\mathbf{P}\{\xi+\eta=l\}}\\&=\mathbf{P}(\eta=k|\xi+\eta=l).\end{aligned}$$

从而 (26) 的第一个等式得证. 为证明第二个等式, 只需注意到, 由于 ξ 和 η 同分布,

$$2\mathbf{E}(\xi|\xi+\eta)=\mathbf{E}(\xi|\xi+\eta)+\mathbf{E}(\eta|\xi+\eta)=\mathbf{E}(\xi+\eta|\xi+\eta)=\xi+\eta.$$

3. 关于 $\mathbf{E}(\xi|\mathscr{B})$ 和 $\mathbf{E}(\xi|\mathscr{D})$ 在 §1 曾指出, 有限集合 Ω 的每一个分割 $\mathscr{D}=\{D_1,\cdots,D_k\}$ 对应着 Ω 子集的一个代数 $\alpha(\mathscr{D})$. 恰好相反, 有限空间 Ω 的任何代数 $\mathscr{B}$ 可以由某一个分割 $\mathscr{D}$ 派生: $\mathscr{B}=\alpha(\mathscr{D})$. 因此在有限空间 Ω 的任何代数与分割之间存在一一对应关系. 这一事实是指, 以后将引进的关于特殊集系 ($\sigma-$ 代数) 的条件数学期望的概念.

对于有限空间, 代数和 $\sigma-$ 代数的概念完全相同. 在这种情形下, 如果 $\mathscr{B}$ 是代数, 则以后 (在第二章 §7) 引进的随机变量 ξ 关于代数 $\mathscr{B}$ 的条件数学期望 $\mathbf{E}(\xi|\mathscr{B})$, 干脆与 $\mathbf{E}(\xi|\mathscr{D})$ 一致, 其中 $\mathbf{E}(\xi|\mathscr{D})$ 是 ξ 关于分割 $\mathscr{D}$ (即关于 $\sigma-$ 代数 $\mathscr{B}=\alpha(\mathscr{D})$) 的条件数学期望. 实际上, 对于有限空间, 我们以后不再区分 $\mathbf{E}(\xi|\mathscr{B})$ 和 $\mathbf{E}(\xi|\mathscr{D})$: 认为 $\mathbf{E}(\xi|\mathscr{B})$ 按定义就是 $\mathbf{E}(\xi|\mathscr{D})$.

4. 练习题

1. 试举一例: 随机变量 ξ 和 η 不独立, 但是

$$\mathbf{E}(\xi|\eta)=\mathbf{E}\xi.$$

(与命题 (22) 对照.)

2. 随机变量

$$\mathbf{D}(\xi|\mathscr{D})=\mathbf{E}\{[\xi-\mathbf{E}(\xi|\mathscr{D})]^2|\mathscr{D}\}$$

称做关于分割 $\mathscr{D}$ 的*条件方差*.

证明方差

$$\mathbf{D}\xi=\mathbf{E}\mathbf{D}(\xi|\mathscr{D})+\mathbf{D}\mathbf{E}(\xi|\mathscr{D}).$$

3. 基于 (17) 式, 证明: 对于任意函数 $f=f(\eta)$, 条件数学期望具有如下性质:

$$\mathbf{E}[f(\eta)\mathbf{E}(\xi|\eta)]=\mathbf{E}[\xi f(\eta)].$$

4. 设 ξ 和 η 是随机变量. 证明

$$\inf_f \mathbf{E}[\eta - f(\xi)]^2.$$

在函数 $f^*(\xi) = \mathbf{E}(\eta|\xi)$ 上达到下确界, 即条件数学期望 $\mathbf{E}(\eta|\xi)$, 在均方意义下是 η 对 ξ 的最优估计量.

5. 设 $\xi_1, \cdots, \xi_k, \tau$ 是独立随机变量, 而 $\xi_1, \cdots, \xi_k$ 同分布, 且 τ 的可能值为 $1, 2, \cdots, n$. 证明, 对于随机个随机变量之和 $S_\tau = \xi_1 + \cdots + \xi_\tau$, 有

$$\mathbf{E}(S_\tau|\tau) = \tau\mathbf{E}\xi_1, \quad \mathbf{D}(S_\tau|\tau) = \tau\mathbf{D}\xi_1$$

和

$$\mathbf{E}S_\tau = \mathbf{E}\tau \times \mathbf{E}\xi_1, \quad \mathbf{D}S_\tau = \mathbf{E}\tau \times \mathbf{D}\xi_1 + \mathbf{D}\tau \times (\mathbf{E}\xi_1)^2.$$

6. 证明等式 (24).

§9. 随机游动 I. 掷硬币博弈的破产概率和平均持续时间

1. 关于随机游动 在 §6 中证明的伯努利概型的极限定理, 远远不只是给出了计算概率 $\mathbf{P}\{S_n = k\}$ 和 $\mathbf{P}\{A < S_n \leqslant B\}$ 的方便公式. 其意义还在于, 这些定理具有适合多方面应用的特点, 也就是说, 这些定理不仅适用于只有两个可能值的伯努利随机变量 $\xi_1, \xi_2, \cdots$, 而且适用于更为一般来源的随机变量. 因此, 伯努利概型作为一种简单的模型, 可以导出很多更为一般模型所固有的许多概率规律性.

在这一节和下一节, 将研究一系列新的概率规律性, 有些甚至带有非常不可预测的特点. 全部研究仍然局限于由伯努利概型所描绘的随机游动, 不过许多结论对于更一般情形的随机游动仍然成立.

2. 二人博弈和破产概率 考虑伯努利概型: $(\Omega, \mathscr{A}, \mathbf{P})$, 其中

$$\Omega = \{\omega : \omega = (x_1, \cdots, x_n), x_i = \pm 1\}, \mathscr{A} = \{A : A \subseteq \Omega\}$$
$$\mathbf{P}(\{\omega\}) = p^{\nu(\omega)} q^{n-\nu(\omega)}, \quad \nu(\omega) = \frac{\sum x_i + n}{2},$$

其中 $\mathscr{A}$ 是 Ω 的一切子集的集系. 设 $\xi_i(\omega) = x_i (i = 1, \cdots, n)$, 则 $\xi_1, \cdots, \xi_n$ 是独立伯努利随机变量序列:

$$\mathbf{P}\{\xi_i = 1\} = p, \quad \mathbf{P}\{\xi_i = -1\} = q, \quad p + q = 1.$$

设 $S_0 = 0, S_k = \xi_1 + \cdots + \xi_k, 1 \leqslant k \leqslant n$. 序列 $(S_k)_{k \leqslant n}$ 可以视为由 0 出发的某"质点"随机游动的轨道. 这时 $S_{k+1} = S_k + \xi_{k+1}$, 即如果质点于 k 时在点 S_k, 则于 $k+1$ 时质点或者以概率 p 向上移动一步, 或者以概率 q 向下移动一步.

设 A 和 $B (A \leqslant 0 \leqslant B)$ 是两个整数. 与所考虑的随机游动相联系的有趣的问题之一就是, 随机游动的质点经 n 步越出区间 (A, B) 的概率如何? 同样有趣的问题是, 经 n 步在点 A 或点 B 越出区间 (A, B) 的概率如何?

如果用博弈来说明, 则该问题的自然性就显得格外清楚. 假设有两人 (甲和乙) 对弈, 开始他们的 "赌金" 相等, 分别为 $(-A)$ 和 B. 若 $\xi_i=+1$, 则认为乙付给甲一单位赌金; 若 $\xi_i=-1$, 则相反甲付给乙一单位赌金. 那么, $S_k=\xi_1+\cdots+\xi_k$ 表示经 k 局甲赢得乙的数额 ($S_k<0$ 实际上表示甲输给乙的数额).

在 $k\leqslant n$ 时, 即当首次 $S_k=B(S_k=A)$ 时, 乙 (甲) 的赌金成为 0, 即出现乙 (甲) *破产*. (如果 $k<n$, 则应该认为在时刻 k 对弈停止, 尽管在 n 时前包括 n 时游动是确定的.)

在给出问题的确切表述之前, 首先引进一系列记号.

设 x 是区间 $[A,B]$ 上的整数, 而对于 $0\leqslant k\leqslant n$, 设 $S_k^x=x+S_k$

$$\tau_k^x=\min\{0\leqslant l\leqslant k:S_l^x=A\text{或}B\},\tag{1}$$

其中约定: 如果对于一切 $0\leqslant l\leqslant k$, 有 $A<S_l^x<B$, 则认为 $\tau_k^x=k$.

对于每一个 $0\leqslant k\leqslant n$ 和 $x\in[A,B]$, 时刻 τ_k^x 称做*停止时间* (见 §11), 是一定义在基本事件空间 Ω 上的整数值随机变量 (注意, 没有明显地标出 τ_k^x 依赖于 ω).

显然, 对于一切 $l<k$, 集合 $\{\omega:\tau_k^x=l\}$ 是一事件: "于 0 时始于点 x 的随机游动 $\{S_i^x:0\leqslant i\leqslant k\}$, 于 l 时越出区间 (A,B)". 同样显然, 对于 $l<k$, 集合 $\{\omega:\tau_k^x=l,S_l^x=A\}$ 与 $\{\omega:\tau_k^x=l,S_l^x=B\}$ 是事件 "游动的质点于 l 时相应地在点 A 和 B 越出区间 (A,B)".

对于一切 $0\leqslant k\leqslant n$, 设

$$\mathscr{A}_k^x=\sum_{0\leqslant l\leqslant k}\{\omega:\tau_k^x=l,S_l^x=A\},\tag{2}$$

$$\mathscr{B}_k^x=\sum_{0\leqslant l\leqslant k}\{\omega:\tau_k^x=l,S_l^x=B\};$$

且设

$$\alpha_k(x)=\mathbf{P}(\mathscr{A}_k^x),\quad \beta_k(x)=\mathbf{P}(\mathscr{B}_k^x)$$

是经时间段 $[0,k]$ 分别于点 A 和 B 处越出区间 (A,B) 的概率. 对于这些概率可以得到递推关系式, 并由此依次求出 $\alpha_1(x),\cdots,\alpha_n(x)$ 和 $\beta_1(x),\cdots,\beta_n(x)$.

这样, 设 $A<x<B$. 显然 $\alpha_0(x)=\beta_0(x)=0$. 设 $1\leqslant k\leqslant n$, 那么由 §3,(3) 式, 有

$$\begin{aligned}\beta_k(x)&=\mathbf{P}(\mathscr{B}_k^x)\\&=\mathbf{P}(\mathscr{B}_k^x|S_1^x=x+1)\mathbf{P}\{\xi_1=1\}+\mathbf{P}(\mathscr{B}_k^x|S_1^x=x-1)\mathbf{P}\{\xi_1=-1\}\\&=p\mathbf{P}(\mathscr{B}_k^x|S_1^x=x+1)+q\mathbf{P}(\mathscr{B}_k^x|S_1^x=x-1).\end{aligned}\tag{3}$$

现在证明

$$\mathbf{P}(\mathscr{B}_k^x|S_1^x=x+1)=\mathbf{P}(\mathscr{B}_{k-1}^{x+1}),\quad \mathbf{P}(\mathscr{B}_k^x|S_1^x=x-1)=\mathbf{P}(\mathscr{B}_{k-1}^{x-1}).$$

为此注意到, 集合 $\mathscr{B}_k^x$ 可以表示为

$$\mathscr{B}_k^x = \{\omega : (x, x+\xi_1, \cdots, x+\xi_1+\cdots+\xi_k) \in B_k^x\},$$

其中 B_k^x 是形如

$$(x, x+x_1, \cdots, x+x_1+\cdots+x_k), \quad x_i = \pm 1$$

轨道的集合, 这样的轨道是在时段 $[0,k]$ 内, 在点 B 首次越出区间 (A,B) (见图 15).

把集合 B_k^x 表示为 $B_k^{x,x+1} + B_k^{x,x-1}$, 其中 $B_k^{x,x+1}$ 和 $B_k^{x,x-1}$ 是 B_k^x 中分别对应于 $x_1 = +1$ 和 $x_1 = -1$ 轨道.

易见, $B_k^{x,x+1}$ 中的每一条轨道与轨道

$$(x, x+1, x+1+x_2, \cdots, x+1+x_2+\cdots+x_k),$$

一一对应. 对于 $B_k^{x,x-1}$ 的轨道, 也有同样的性质. 由于这一性质, 以及随机变量 $\xi_1, \cdots, \xi_k$ 独立同分布和 §8 式 (6), 有

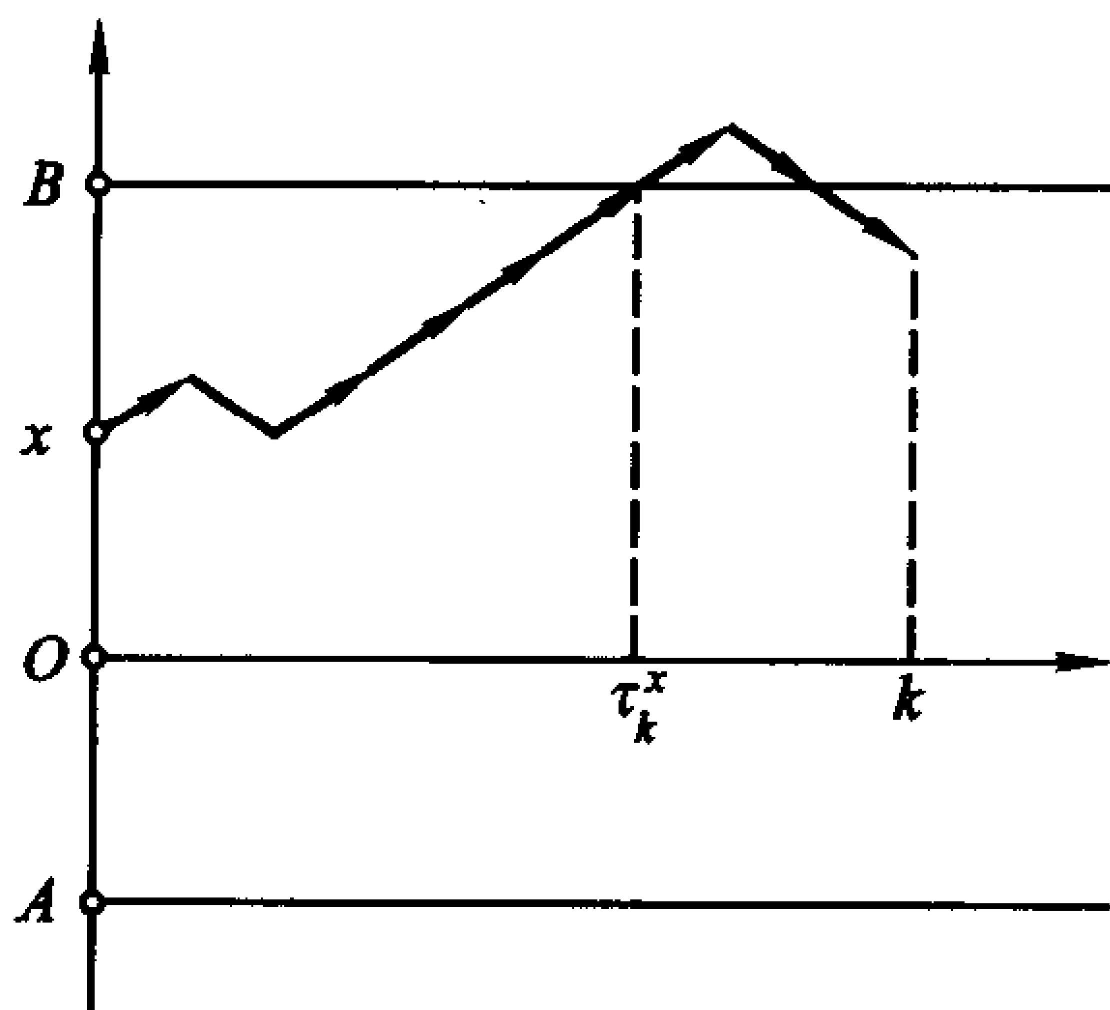


图 15 B_k^x 轨道的例

$$\begin{aligned}
&\mathbf{P}(\mathscr{B}_k^x | S_1^x = x+1) = \mathbf{P}(\mathscr{B}_k^x | \xi_1 = 1) \\
&= \mathbf{P}\{(x, x+\xi_1, \cdots, x+\xi_1+\cdots+\xi_k) \in B_k^x | \xi_1 = 1\} \\
&= \mathbf{P}\{(x+1, x+1+\xi_2, \cdots, x+1+\xi_2+\cdots+\xi_k) \in B_{k-1}^{x+1}\} \\
&= \mathbf{P}\{(x+1, x+1+\xi_1, \cdots, x+1+\xi_1+\cdots+\xi_{k-1}) \in B_{k-1}^{x+1}\} \\
&= \mathbf{P}(\mathscr{B}_{k-1}^{x+1}).
\end{aligned}$$

同样地

$$\mathbf{P}(\mathscr{B}_k^x | S_1^x = x-1) = \mathbf{P}(\mathscr{B}_{k-1}^{x-1}).$$

因此, 对于 $x \in (A,B)$ 和 $k \leqslant n$, 由 (3) 式, 有

$$\beta_k(x) = p\beta_{k-1}(x+1) + q\beta_{k-1}(x-1), \tag{4}$$

其中

$$\beta_l(B) = 1, \quad \beta_l(A) = 0, \quad 0 \leqslant l \leqslant n. \tag{5}$$

同理

$$\alpha_k(x) = p\alpha_{k-1}(x+1) + q\alpha_{k-1}(x-1), \tag{6}$$

其中

$$\alpha_l(A) = 1, \quad \alpha_l(B) = 0, \quad 0 \leqslant l \leqslant n.$$

由于 $\alpha_0(x)=\beta_0(x)=0, x\in(A,B)$, 则所得递推公式可以 (至少原则上可以) 用来求概率 $\alpha_1(x),\cdots,\alpha_n(x)$ 和 $\beta_1(x),\cdots,\beta_n(x)$. 暂时不作具体计算, 而考虑当 n 充分大时这些概率的值的问题.

为此注意到, 由于 $\mathscr{B}_{k-1}^x\subset\mathscr{B}_k^x, k\leqslant n$, 可见 $\beta_{k-1}(x)\leqslant\beta_k(x)\leqslant 1$. 因此自然想到 (实际上确实是这样, 见下面第 3 小节), 当 n 充分大时概率 $\beta_n(x)$ 接近如下方程的解 $\beta(x)$:

$$\beta(x)=p\beta(x+1)+q\beta(x-1), \tag{7}$$

且满足由 (4) 和 (5) 两式求极限所得的边界条件:

$$\beta(B)=1,\ \beta(A)=0. \tag{8}$$

为求解问题 (7) 和 (8), 首先假设 $p\neq q$. 不难看到, 所考虑的方程有两个特解: a 和 $b(q/p)^x$, 其中 a 和 b 是常数. 因此, 我们以如下形式求通解 $\beta(x)$:

$$\beta(x)=a+b(q/p)^x. \tag{9}$$

考虑到 (8) 式, 对于一切 $A\leqslant x\leqslant B$, 有

$$\beta(x)=\frac{(q/p)^x-(q/p)^A}{(q/p)^B-(q/p)^A}. \tag{10}$$

现在证明这是该问题的唯一解. 为此, 只需证明, 问题 (7) 和 (8) 式的一切解都可以表示为 (9) 的形式.

设 $\widetilde{\beta}(x)$ 是满足 $\widetilde{\beta}(A)=0,\widetilde{\beta}(B)=1$ 的某个解. 总存在常数 $\widetilde{a}$ 和 $\widetilde{b}$, 使

$$\widetilde{a}+\widetilde{b}(q/p)^A=\widetilde{\beta}(A),\quad \widetilde{a}+\widetilde{b}(q/p)^{A+1}=\widetilde{\beta}(A+1).$$

因此由方程 (7) 可见

$$\widetilde{\beta}(A+2)=\widetilde{a}+\widetilde{b}(q/p)^{A+2},$$

且一般

$$\widetilde{\beta}(x)=\widetilde{a}+\widetilde{b}(q/p)^x.$$

于是, 所求得的解 (10) 是所考虑问题的唯一解.

由类似的讨论可见, 方程

$$\alpha(x)=p\alpha(x+1)+q\alpha(x-1),\quad x\in(A,B) \tag{11}$$

满足边界条件

$$\alpha(A)=1,\quad \alpha(B)=0 \tag{12}$$

的唯一解是

$$\alpha(x)=\frac{(q/p)^B-(q/p)^x}{(q/p)^B-(q/p)^A},\quad A\leqslant x\leqslant B. \tag{13}$$

假如 $p=q=1/2$, 则问题 (7),(8) 和 (11),(12) 唯一解 $\beta(x)$ 和 $\alpha(x)$ 相应为:

$$\beta(x)=\frac{x-A}{B-A} \tag{14}$$

和

$$\alpha(x)=\frac{B-x}{B-A}. \tag{15}$$

注意, 对于任意 $0 \leqslant p \leqslant 1$,

$$\alpha(x)+\beta(x)=1. \tag{16}$$

假如对局二人甲和乙的最初的 "赌金" 分别为 $x-A$ 和 $B-x$, 对于无限多局的对弈, 宜把 $\alpha(x)$ 和 $\beta(x)$ 相应地称做对局二人**甲和乙的破产概率**, 当然应假定存在**无限**独立伯努利随机变量序列 $\xi_1, \xi_2, \cdots$, 其中 $\xi_i=+1$ 表示甲赢, 而 $\xi_i=-1$ 表示甲输. 为在本小节开始考虑的概率空间 $(\Omega, \mathscr{A}, \mathbf{P})$ 上, 存在上述独立伯努利随机变量序列, $(\Omega, \mathscr{A}, \mathbf{P})$ 显得太 "贫乏". 以后 (第二章 §9) 我们会看到, **确实**可以建立这样的**无限**序列, 而且 $\alpha(x)$ 和 $\beta(x)$ 确实是在无限多步破产的概率.

现在讨论由这些公式得到的一系列推论.

如果设 $A=0, 0 \leqslant x \leqslant B$, 则函数 $\beta(x)$ 按其本来含义, 是自状态 x 出发的质点在到达 0 点之前到达点 B 的概率. 由 (10) 和 (14) 式, 可见 (图 16):

$$\beta(x)=\begin{cases} x/B, & 若\ p=q=1/2, \\ \dfrac{(q/p)^x-1}{(q/p)^B-1}, & 若\ p \neq q. \end{cases} \tag{17}$$

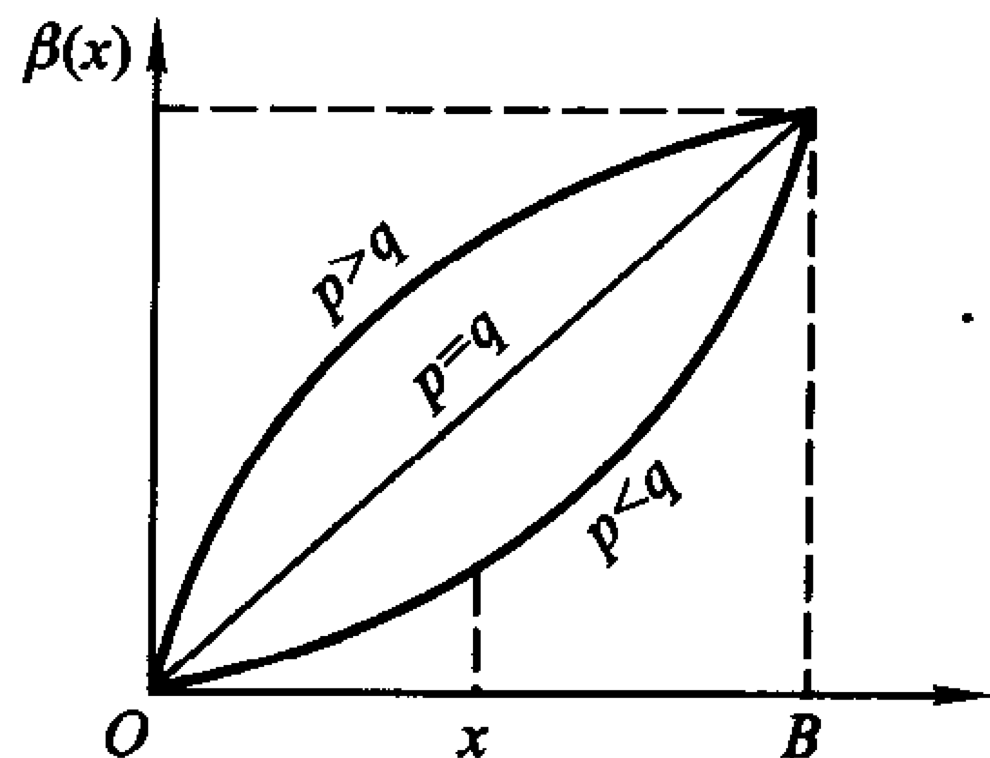


图 16 $\beta(x)$ 的图形

[$\beta(x)$ 是当质点 B 自点 x 出发早于点 0 到达点 B 的概率]

其次, 假设不等式 $q>p$ 成立, 即对于甲是不利博弈, 其破产的极限概率 $\alpha(0)$ 为:

$$\alpha=\frac{(q/p)^B-1}{(q/p)^B-(q/p)^A}.$$

现在假设对弈条件有如下变化: 甲和乙的赌资仍然是 $(-A)$ 和 B, 但是现在每人支付的金额为 1/2 单位, 而不是以前的 1 单位. 换句话说, 现在假设 $\mathbf{P}\{\xi_i=1/2\}=p, \mathbf{P}\{\xi_i=-1/2\}=q$. 这时, 若以 $\alpha_{1/2}$ 表示甲破产的概率, 则

$$\alpha_{1/2}=\frac{(q/p)^{2B}-1}{(q/p)^{2B}-(q/p)^{2A}},$$

因此, 如果 $q>p$, 则

$$\alpha_{1/2}=\alpha \times \frac{(q/p)^B+1}{(q/p)^B+(q/p)^A}>\alpha.$$

由此得出结论: **如果对弈对于甲不利 (即 $q>p$), 则增加一倍赌金可以减小他破产的概率.**

3. $\alpha_n(x)$ 和 $\beta_n(x)$ 收敛于 $\alpha(x)$ 和 $\beta(x)$ 的速度 现在讨论 $\alpha_n(x)$ 和 $\beta_n(x)$ 收敛于极限值 $\alpha(x)$ 和 $\beta(x)$ 的速度问题. 为简便计, 假设 $x=0$, 记

$$\alpha_n=\alpha_n(0),\quad \beta_n=\beta_n(0),\quad \gamma_n=1-(\alpha_n+\beta_n).$$

显然,

$$\gamma_n=\mathbf{P}\{A<S_k<B, 0\leqslant k\leqslant n\},$$

其中 $\{A<S_k<B, 0\leqslant k\leqslant n\}$ 就是事件

$$\bigcap_{0\leqslant k\leqslant n}\{A<S_k<B\}.$$

设 $n=rm$, 其中 r 和 m 是整数; 设

$$\begin{aligned}
\zeta_1&=\xi_1+\cdots+\xi_m,\\
\zeta_2&=\xi_{m+1}+\cdots+\xi_{2m},\\
&\cdots\cdots\cdots\cdots\\
\zeta_r&=\xi_{m(r-1)+1}+\cdots+\xi_{rm}.
\end{aligned}$$

那么, 设 $C=|A|+B$, 易见

$$\{A<S_k<B, 1\leqslant k\leqslant rm\}\subseteq\{|\zeta_1|<C,\cdots,|\zeta_r|<C\},$$

因此由随机变量 $\zeta_1,\cdots,\zeta_r$ 独立同分布, 可见

$$\gamma_n\leqslant\mathbf{P}\{|\zeta_1|<C,\cdots,|\zeta_r|<C\}=\prod_{i=1}^{r}\mathbf{P}\{|\zeta_i|<C\}=(\mathbf{P}\{|\zeta_1|<C\})^r. \tag{18}$$

注意到 $\mathbf{D}\zeta_1=m[1-(p-q)^2]$, 因此当 $0<p<1$ 时, 对于充分大的 m, 有

$$\mathbf{P}\{|\zeta_1|<C\}\leqslant\varepsilon_1, \tag{19}$$

其中 $\varepsilon_1<1$, 因为假如 $\mathbf{P}\{|\zeta_1|<C\}=1$, 则 $\mathbf{D}\zeta_1\leqslant C^2$.

假如 $p=0$ 或 1, 则对于充分大的 m, 概率 $\mathbf{P}\{|\zeta_1|<C\}=0$, 从而对于一切 $0\leqslant p\leqslant 1$, (19) 式成立.

由 (18) 和 (19) 两式, 可见对于充分大的 n,

$$\gamma_n\leqslant\varepsilon^n, \tag{20}$$

其中 $\varepsilon=\varepsilon_1^{1/m}<1$.

由 (16) 式, $\alpha+\beta=1$. 因此

$$(\alpha-\alpha_n)+(\beta-\beta_n)=\gamma_n$$

而因为 $\alpha \geqslant \alpha_n, \beta \geqslant \beta_n$, 所以

$$0 \leqslant \alpha - \alpha_n \leqslant \gamma_n \leqslant \varepsilon^n,$$
$$0 \leqslant \beta - \beta_n \leqslant \gamma_n \leqslant \varepsilon^n,$$

其中 $\varepsilon < 1$. 对于 $\alpha(x) - \alpha_n(x)$ 和 $\beta(x) - \beta_n(x)$ 有类似的估计.

4. 平均游动时间 现在讨论平均随机游动时间.

设 $m_k(x) = \mathbf{E}\tau_k^x$ 是停止时间 $\tau_k^x, (k \leqslant n)$ 的数学期望. 仿照推导 $\beta_n(x)$ 的递推公式的方法, 对于 $x \in (A, B)$, 得

$$\begin{aligned}
m_k(x) = \mathbf{E}\tau_k^x &= \sum_{1 \leqslant l \leqslant k} l\mathbf{P}\{\tau_k^x = l\} \\
&= \sum_{1 \leqslant l \leqslant k} l[p\mathbf{P}\{\tau_k^x = l | \xi_1 = 1\} + q\mathbf{P}\{\tau_k^x = l | \xi_1 = -1\}] \\
&= \sum_{1 \leqslant l \leqslant k} l[p\mathbf{P}\{\tau_{k-1}^{x+1} = l - 1\} + q\mathbf{P}\{\tau_{k-1}^{x-1} = l - 1\}] \\
&= \sum_{0 \leqslant l \leqslant k-1} (l+1)[p\mathbf{P}\{\tau_{k-1}^{x+1} = l\} + q\mathbf{P}\{\tau_{k-1}^{x-1} = l\}] \\
&= pm_{k-1}(x+1) + qm_{k-1}(x-1) \\
&\qquad + \sum_{0 \leqslant l \leqslant k-1} [p\mathbf{P}\{\tau_{k-1}^{x+1} = l\} + q\mathbf{P}\{\tau_{k-1}^{x-1} = l\}] \\
&= pm_{k-1}(x+1) + qm_{k-1}(x-1) + 1.
\end{aligned}$$

于是, 对于 $x \in (A, B)$ 和 $0 \leqslant k \leqslant n$, 函数 $m_k(x)$ 满足递推关系式:

$$m_k(x) = 1 + pm_{k-1}(x+1) + qm_{k-1}(x-1), \tag{21}$$

其中 $m_0(x) = 0$. 由这些方程连同边界条件

$$m_k(A) = m_k(B) = 0, \tag{22}$$

可以依次求出 $m_1(x), \cdots, m_n(x)$.

由于 $m_k(x) \leqslant m_{k+1}(x)$, 可见存在极限

$$m(x) = \lim_{n \to \infty} m_n(x),$$

且由 (21) 式知 $m(x)$ 满足方程

$$m(x) = 1 + pm(x+1) + qm(x-1) \tag{23}$$

和边界条件

$$m(A) = m(B) = 0. \tag{24}$$

为求该方程的解, 首先假设

$$m(x) < \infty, x \in (A, B). \tag{25}$$

那么, 如果 $p \neq q$, 则特解的形式为 $x/(q-p)$, 而通解 (见 (9) 式) 可以表示为:

$$m(x) = \frac{x}{p-q} + a + b\left(\frac{q}{p}\right)^x.$$

由此并考虑到边界条件 $m(A) = m(B) = 0$, 即可求出

$$m(x) = \frac{1}{p-q}[B\beta(x) + A\alpha(x) - x], \tag{26}$$

其中 $\beta(x)$ 和 $\alpha(x)$ 由 (10) 和 (13) 两式决定. 如果 $p = q = 1/2$, 则方程 (23) 的通解形如

$$m(x) = a + bx - x^2,$$

因此, 由于边界条件 $m(A) = m(B) = 0$, 可见

$$m(x) = (B - x)(x - A). \tag{27}$$

特别, 由此可见, 如果对弈二人的初始赌金相等: $B = -A$, 则

$$m(0) = B^2.$$

假设 $B = 10$ 且每秒一局, 那么某一对手破产前的 (极限) 平均时间相当长: 等于 100 秒.

公式 (26) 和 (27) 是在假设 $m(x) < \infty, x \in (A, B)$ 的条件下得到的, 现在证明, 实际上 $m(x)$ 对于一切 $x \in (A, B)$ 有限. 我们只限于讨论 $x = 0$ 的情形. 一般情形可以类似地证明.

设 $p = q = 1/2$. 将数列 $S_0, S_1, \cdots, S_n$ 与停止时间 $\tau_n = \tau_n^0$, 由随机变量 $S_{\tau_n} = S_{\tau_n}(\omega)$ 联系在一起, 其中

$$S_{\tau_n}(\omega) = \sum_{k=0}^{n} S_k(\omega) I_{\{\tau_n = k\}}(\omega). \tag{28}$$

随机变量 S_{τ_n} 的直观意义是清楚的: 它是随机游动在停止时刻 τ_n 的值. 这时, 假如 $\tau_n < n$, 则 $S_{\tau_n} = A$ 或 B; 而如果 $\tau_n = n$, 则 $A \leqslant S_{\tau_n} \leqslant B$.

现在证明, 对于 $p = q = 1/2$, 有

$$\mathbf{E} S_{\tau_n} = 0, \tag{29}$$

$$\mathbf{E} S_{\tau_n}^2 = \mathbf{E}\tau_n. \tag{30}$$

为证明 (29) 式, 注意到

$$
\begin{aligned}
\mathbf{E}S_{\tau_n}^2 &= \sum_{k=0}^{n} \mathbf{E}[S_k I_{\{\tau_n=k\}}(\omega)] \\
&= \sum_{k=0}^{n} \mathbf{E}[S_n I_{\{\tau_n=k\}}(\omega)] + \sum_{k=0}^{n} \mathbf{E}[(S_k - S_n) I_{\{\tau_n=k\}}(\omega)] \\
&= \mathbf{E}S_n + \sum_{k=0}^{n} \mathbf{E}[(S_k - S_n) I_{\{\tau_n=k\}}(\omega)],
\end{aligned} \tag{31}
$$

其中显然 $\mathbf{E}S_n = 0$. 现在证明

$$
\sum_{k=0}^{n} \mathbf{E}[(S_k - S_n) I_{\{\tau_n=k\}}(\omega)] = 0.
$$

对于 $0 \leqslant k < n$, 有 $\{\tau_n > k\} = \{A < S_1 < B, \cdots, A < S_k < B\}$. 事件

$$
\{\omega : A < S_1 < B, \cdots, A < S_k < B\}
$$

显然可以表示为

$$
\{\omega : (\xi_1, \cdots, \xi_k) \in A_k\}, \tag{32}
$$

其中 A_k 是集合 $\{-1, +1\}^k$ 的某一子集. 换句话说, 这是一个只决定于随机变量 $\xi_1, \cdots, \xi_k$ 的值, 而不依赖于随机变量 $\xi_{k+1}, \cdots, \xi_n$ 值的集合. 由于集合

$$
\{\tau_n = k\} = \{\tau_n > k-1\} \backslash \{\tau_n > k\},
$$

可见它也是形如 (32) 的集合. 由于随机变量 $\xi_1, \cdots, \xi_k$ 的独立, 以及由 §4 的练习题 10 可见, 对于任意 $0 \leqslant k < n$, 随机变量 $S_n - S_k$ 和 $I_{\{\tau_n=k\}}$ 独立, 从而

$$
\mathbf{E}[(S_n - S_k) I_{\{\tau_n=k\}}] + \mathbf{E}[S_n - S_k] \times \mathbf{E}I_{\{\tau_n=k\}} = 0.
$$

于是 (29) 式得证.

同样可以证明 (30) 式:

$$
\begin{aligned}
\mathbf{E}S_{\tau_n}^2 &= \sum_{k=0}^{n} \mathbf{E}S_k^2 I_{\{\tau_n=k\}} = \sum_{k=0}^{n} \mathbf{E}\{[S_n + (S_k - S_n)]^2 I_{\{\tau_n=k\}}\} \\
&= \sum_{k=0}^{n} [\mathbf{E}S_k^2 I_{\{\tau_n=k\}} + 2\mathbf{E}S_n (S_k - S_n) I_{\{\tau_n=k\}} + \mathbf{E}(S_n - S_k)^2 I_{\{\tau_n=k\}}] \\
&= \mathbf{E}S_n^2 - \sum_{k=0}^{n} \mathbf{E}(S_n - S_k)^2 I_{\{\tau_n=k\}} = n - \sum_{k=0}^{n} (n-k)\mathbf{P}\{\tau_n = k\} \\
&= \sum_{k=0}^{n} k\mathbf{P}\{\tau_n = k\} = \mathbf{E}\tau_n.
\end{aligned}
$$

于是, 对于 $p=q=1/2$, 公式 (29) 和 (30) 得证. 对于任意 $p,q(p+q=1)$ 的情形, 类似地可以证明:

$$\mathbf{E}S_{\tau_n}=(p-q)\mathbf{E}\tau_n, \tag{33}$$

$$\mathbf{E}[S_{\tau_n}-\tau_n\mathbf{E}\xi_1]^2=\mathbf{D}\xi_1\times\mathbf{E}\tau_n, \tag{34}$$

其中 $\mathbf{E}\xi_1=p-q,\mathbf{D}\xi_1=1-(p-q)^2$.

现在利用得到的关系式, 证明

$$\lim_{n\to\infty}m_n(0)=m(0)<\infty.$$

对于 $p=q=1/2$, 由 (30) 式, 有

$$\mathbf{E}\tau_n\leqslant\max\{A^2,B^2\}. \tag{35}$$

如果 $p\neq q$, 则由 (33) 式, 有

$$\mathbf{E}\tau_n\leqslant\frac{\max\{|A|,B\}}{|p-q|}, \tag{36}$$

由此可见 $m(0)<\infty$.

还要指出, 对于 $p=q=1/2$, 有

$$\mathbf{E}\tau_n=\mathbf{E}S_{\tau_n}^2=A^2\alpha_n+B^2\beta_n+\mathbf{E}[S_n^2I_{\{A<S_n<B\}}I_{\{\tau_n=n\}}],$$

因此

$$A^2\alpha_n+B^2\beta_n\leqslant\mathbf{E}\tau_n\leqslant A^2\alpha_n+B^2\beta_n+\max\{A^2,B^2\}\times\gamma_n.$$

由此及不等式 (20) 可见, 当 $n\to\infty$ 时数学期望 $\mathbf{E}\tau_n$ 以指数的速度收敛于极限值:

$$m(0)=A^2\alpha+B^2\beta=A^2\times\frac{B}{B-A}-B^2\times\frac{A}{B-A}=|AB|.$$

对于 $p\neq q$, 有类似的结果 —— $\mathbf{E}\tau_n$ 以指数的速度收敛于极限值:

$$\mathbf{E}\tau_n\to m(0)=\frac{\alpha A+\beta B}{p-q}.$$

5. 练习题

1. 证明 (33) 和 (34) 式的推广, 下列公式成立:

$$\mathbf{E}S_{\tau_n^x}^x=x+(p-q)\mathbf{E}\tau_n^x,$$

$$\mathbf{E}[S_{\tau_n^x}^x-\tau_n^x\mathbf{E}\xi_1]^2=\mathbf{D}\xi_1\times\mathbf{E}\tau_n^x+x^2.$$

2. 讨论当水平 $A\downarrow-\infty$ 时, $\alpha(x),\beta(x)$ 和 $m(x)$ 趋向何量的问题.

3. 假设伯努利概型中 $p=q=1/2$, 问当 n 充分大时 $\mathbf{E}|S_n|$ 是何数量级?

4. 甲乙二人各自独立地掷一枚对称的硬币. 证明, 他们各自掷 n 次, 二人掷出正面的次数相同的概率为

$$\frac{1}{2^{2n}}\sum_{k=0}^{n}(\mathrm{C}_n^k)^2,$$

并由此导出恒等式 (亦见 §2 练习题 4)

$$\sum_{k=0}^{n}(\mathrm{C}_n^k)^2=\mathrm{C}_{2n}^n.$$

设 σ_n 表示时刻: 对手甲和乙首次各自掷出正面的次数相等 (若这样的时刻不存在, 则令 $\sigma_n=n+1$). 求数学期望 $\mathbf{E}\min\{\sigma_n,n\}$.

§10. 随机游动 II. 反射原理. 反正弦定律

1. 质点首次返回 0 的时间　像上一节一样, 假设 $\xi_1,\cdots,\xi_{2n}$ 是独立同分布伯努利随机变量序列, 且

$$\mathbf{P}\{\xi_i=1\}=p,\quad \mathbf{P}\{\xi_i=-1\}=q,$$
$$S_k=\xi_1+\cdots+\xi_k,\quad 1\leqslant k\leqslant 2n\quad S_0=0.$$

记

$$\sigma_{2n}=\min\{1\leqslant k\leqslant 2n:S_k=0\},$$

对于一切 $1\leqslant k\leqslant 2n$, 若 $S_k\neq 0$, 则设 $\sigma_{2n}=\infty$.

量 σ_{2n} 的直观含义十分清楚: 它是*首次返回* 0 的时刻. 这一节将要研究首次返回 0 的时刻 σ_{2n} 的性质, 这时假设所考虑的随机游动对称, 即 $p=q=1/2$.

对于 $0\leqslant k\leqslant 2n$, 记

$$u_{2k}=\mathbf{P}\{S_{2k}=0\},\quad f_{2k}=\mathbf{P}\{\sigma_{2n}=2k\}.\tag{1}$$

显然, $u_0=1$ 而

$$u_{2k}=\mathrm{C}_{2k}^k\frac{1}{2^{2k}}.$$

我们最近的目标是证明, 对于 $1\leqslant k\leqslant n$, 概率决定于公式:

$$f_{2k}=\frac{1}{2k}u_{2(k-1)}.\tag{2}$$

易见, 对于 $1\leqslant k\leqslant n$, 有

$$\{\sigma_{2n}=2k\}=\{S_1\neq 0,\cdots,S_{2k-1}\neq 0,S_{2k}=0\},$$

由对称性, 可见

$$f_{2k} = \mathbf{P}(S_1 \neq 0, \cdots, S_{2k-1} \neq 0, S_{2k} = 0)$$
$$= 2\mathbf{P}(S_1 > 0, \cdots, S_{2k-1} > 0, S_{2k} = 0). \tag{3}$$

我们把序列 $(S_0, \cdots, S_k)$ 称做长为 k 的路径, 以 $L_k(A)$ 表示具有性质 A 的长为 k 的路径条数. 那么,

$$f_{2k} = \frac{1}{2^{2n-1}} \sum_{\{a_{2k+1},\cdots,a_{2n}\}} [L_{2n}(S_1 > 0, \cdots, S_{2k-1} > 0, S_{2k} = 0,$$
$$S_{2k+1} = a_{2k+1}, \cdots, S_{2n} = a_{2k+1} + \cdots + a_{2n})]$$
$$= \frac{1}{2^{2k-1}} L_{2k}(S_1 > 0, \cdots, S_{2k-1} > 0, S_{2k} = 0), \tag{4}$$

其中对一切 $\{a_{2k+1}, \cdots, a_{2n}\}, a_i = \pm 1$ 的组合求和.

于是, 求概率 f_{2k} 归结为求路径的条数 $L_{2k}\{S_1 > 0, \cdots, S_{2k-1} > 0, S_{2k} = 0\}$.

引理 1 设 a, b 为非负整数, $a - b > 0, k = a + b$, 则

$$L_k(S_1 > 0, \cdots, S_{k-1} > 0, S_k = a - b) = \frac{a-b}{k} \mathrm{C}_k^a. \tag{5}$$

证明 事实上

$$L_k(S_1 > 0, \cdots, S_{k-1} > 0, S_k = a - b)$$
$$= L_k(S_1 = 1, S_2 > 0, \cdots, S_{k-1} > 0, S_k = a - b)$$
$$= L_k(S_1 = 1, S_k = a - b)$$
$$- L_k(S_1 = 1, S_k = a - b; \exists i, 2 \leqslant i \leqslant k-1, \text{ 使 } S_i \leqslant 0). \tag{6}$$

换句话说, 始自点 $(1, 1)$ 止于点 $(k, a-b)$ 的**正路径** $(S_1, S_2, \cdots, S_k)$ 的条数, 等于由点 $(1,1)$ 到点 $(k, a-b)$ 的全部路径条数, 减去与时轴相切或相交的路径条数*).

注意到,

$$L_k(S_1 = 1, S_k = a - b; \exists i, 2 \leqslant i \leqslant k-1, \text{ 使 } S_i \leqslant 0) = L_k(S_1 = -1, S_k = a - b), \tag{7}$$

即自点 $\alpha = (1, 1)$ 到点 $\beta = (k, a - b)$ 且与时轴相切或相交的所有路径条数, 等于由点 $\alpha^* = (1, -1)$ 到点 $\beta = (k, a-b)$ 的所有路径的条数. 这一事实称做**反射原理**. 容易证明, 在连接 α 和 β 两点的路径 $A = (S_1, \cdots, S_a, S_{a+1}, \cdots, S_k)$, 与在连接 α^* 和 β 两点的路径 $B = (-S_1, \cdots, -S_a, S_{a+1}, \cdots, S_k)$, 之间存在一一对应关系 (图 17), 其中 a 是路径 A 和路径 B 为 0 的第一个点; 其证明可以由上述两条路径的一一对应关系导出.

*) 路径 $(S_1, S_2, \cdots, S_k)$ 称做**正的** (**非负的**), 如果全部 $S_i > 0$ $(S_i \geqslant 0)$; 路径称做**与时轴相切的**, 如果对于所有 $1 \leqslant j \leqslant k, S_j \geqslant 0$ $(S_j \leqslant 0)$, 且存在 $1 \leqslant i \leqslant k, S_i = 0$; 路径称做**与时轴相交的**, 如果存在两个时点 i 和 j, 使 $S_j > 0$ $(S_j < 0)$.

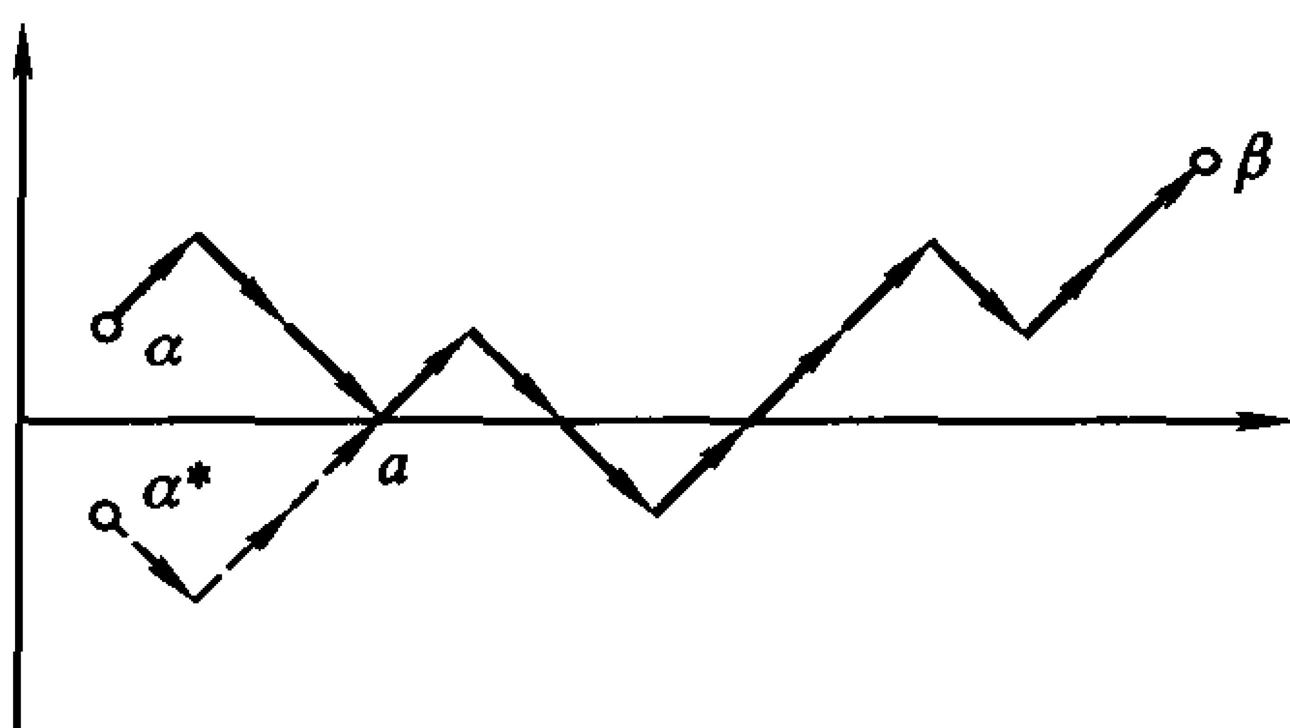


图 17 反射原理

由 (6) 式和 (7) 式可见

$$
\begin{aligned}
& L_k(S_1 > 0, \cdots, S_{k-1} > 0, S_k = a - b) \\
= & L_k(S_1 = 1, S_k = a - b) - L_k(S_1 = -1, S_k = a - b) \\
= & \mathrm{C}_{k-1}^{a-1} - \mathrm{C}_{k-1}^{a} = \frac{a-b}{k}\mathrm{C}_k^a,
\end{aligned}
$$

于是, 命题 (5) 得证. □

现在, 回到概率 f_{2k} 的计算. 由 (4) 式和 (5) 式 ($a = k, b = k-1$), 可见

$$
\begin{aligned}
f_{2k} &= 2^{-2k+1} L_{2k}(S_1 > 0, \cdots, S_{2k-1} > 0, S_{2k} = 0) \\
&= 2^{-2k+1} L_{2k-1}(S_1 > 0, \cdots, S_{2k-1} = 1) \\
&= 2^{-2k+1} \times \frac{1}{2k-1} \times \mathrm{C}_{2k-1}^k = \frac{1}{2k} u_{2(k-1)}
\end{aligned}
$$

于是, 公式 (2) 得证.

我们再给出公式 (2) 的一个证明. 证明基于如下事实:

$$
\frac{1}{2k} u_{2(k-1)} = u_{2(k-1)} - u_{2k}. \tag{8}
$$

显然,

$$
\begin{aligned}
\{\sigma_{2n} = 2k\} &= \{\sigma_{2n} > 2(k-1)\} \backslash \{\sigma_{2n} > 2k\}, \\
\{\sigma_{2n} > 2l\} &= \{S_1 \neq 0, \cdots, S_{2l} \neq 0\},
\end{aligned}
$$

即

$$
\{\sigma_{2n} = 2k\} == \{S_1 \neq 0, \cdots, S_{2(k-1)} \neq 0\} \backslash \{S_1 \neq 0, \cdots, S_{2k} \neq 0\}.
$$

因此

$$
f_{2k} = \mathbf{P}\{S_1 \neq 0, \cdots, S_{2(k-1)} \neq 0\} - \mathbf{P}\{S_1 \neq 0, \cdots, S_{2k} \neq 0\},
$$

从而, 由于 (8) 式, 为证明等式

$$
f_{2k} = \frac{1}{2k} u_{2(k-1)},
$$

只需证明

$$L_{2k}(S_1 \neq 0, \cdots, S_{2k} \neq 0) = L_{2k}(S_{2k} = 0). \tag{9}$$

为此, 注意到, 显然

$$L_{2k}(S_1 \neq 0, \cdots, S_{2k} \neq 0) = 2L_{2k}(S_1 > 0, \cdots, S_{2k} > 0),$$

因此为验证 (9) 式, 只需证明

$$2L_{2k}(S_1 > 0, \cdots, S_{2k} > 0) = L_{2k}(S_1 \geqslant 0, \cdots, S_{2k} \geqslant 0), \tag{10}$$

且

$$L_{2k}(S_1 \geqslant 0, \cdots, S_{2k} \geqslant 0) = L_{2k}(S_{2k} = 0). \tag{11}$$

如果证明在至少一个 $S_i = 0$ 的路径 $A = (S_1, \cdots, S_{2k})$, 与正路径 $B = (S_1, \cdots, S_{2k})$ 之间存在一一对应关系, 则就可以证明等式 (10).

设 $A = (S_1, \cdots, S_{2k})$ 是在点为首次 0 (即 $S_a = 0$) 的非负路径. 自点 $(a, 2)$ 引一轨道 $(S_a + 2, S_{a+1} + 2, \cdots, S_{2k} + 2)$ (图 18 中虚线), 那么路径 $B = (S_1, \cdots, S_{a-1}, S_a + 2, \cdots, S_{2k} + 2)$ 是正的. 相反, 设 $B = (S_1, \cdots, S_{2k})$ 是某一正路径, 而 b 是 $S_b = 1$ 的最后时刻 (图 19). 那么路径 $A = (S_1, \cdots, S_b, S_{b+1} - 2, \cdots, S_k - 2)$ 是非负的. 由上面的构造可见, 在正路径和 (至少有一个 $S_i = 0$) 的非负路径之间, 存在一一对应关系. 于是等式 (10) 得证.

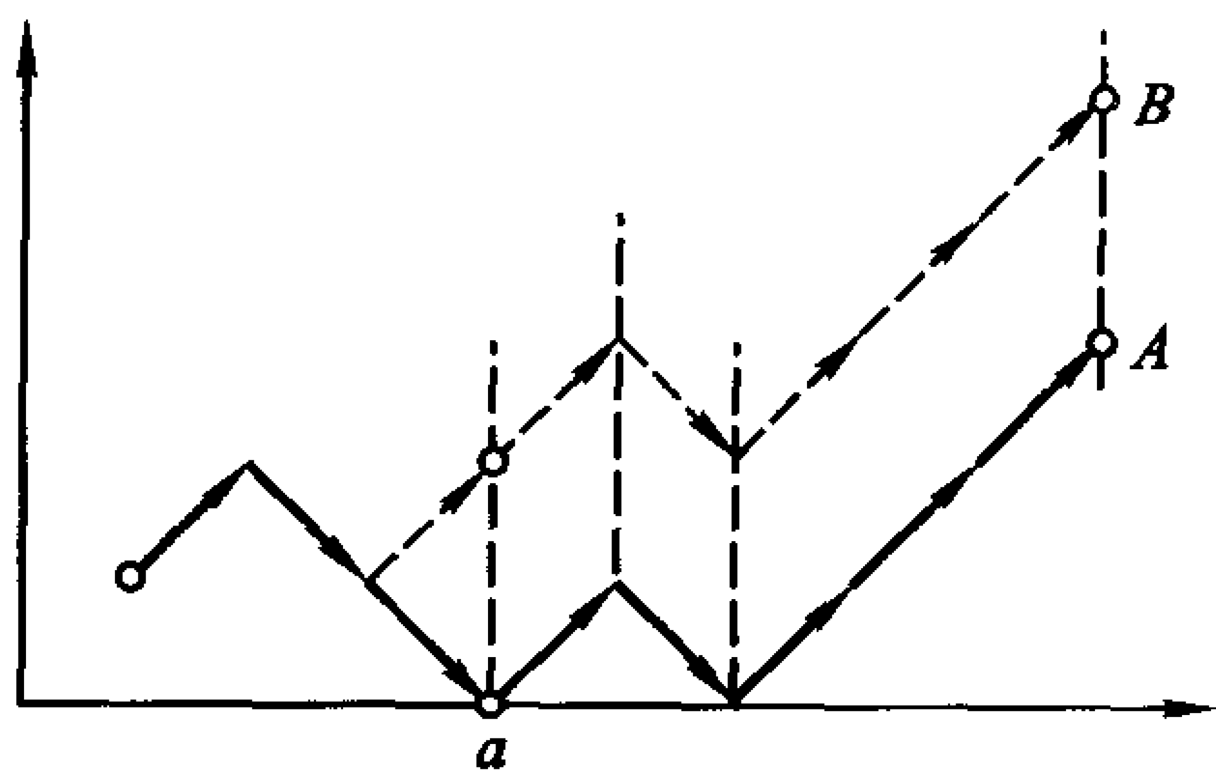


图 18

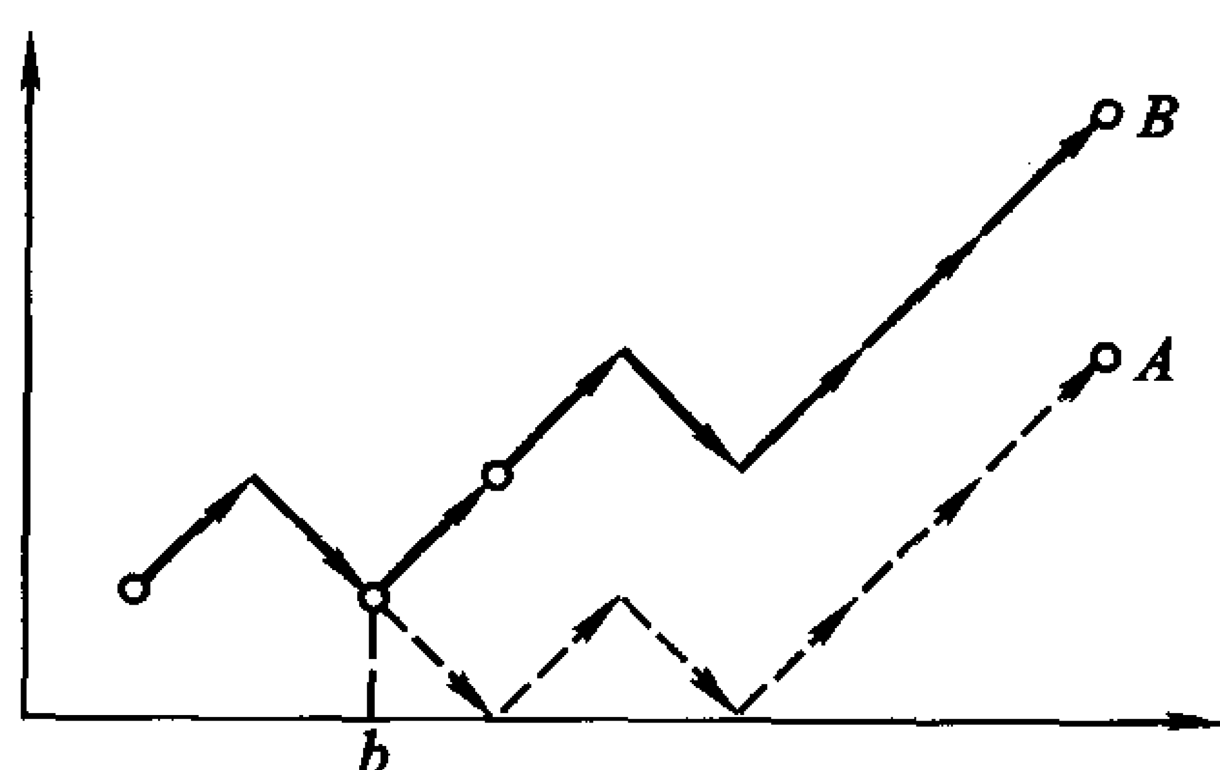


图 19

现在证明等式 (11). 由于对称性和 (10) 式, 只需要证明

$$L_{2k}(S_1 > 0, \cdots, S_{2k} > 0) + L_{2k}(S_1 \geqslant 0, \cdots, S_{2k} \geqslant 0, \text{且 } \exists i, 0 \leqslant i \leqslant 2k, S_i = 0) = L_{2k}(S_{2k} = 0).$$

路径的集合 $(S_{2k} = 0)$ 可以表示为两个集合 $\mathscr{C}_1$ 与 $\mathscr{C}_2$ 之和, 其中 $\mathscr{C}_1$ 是只有一个极小值的路径 $(S_0, \cdots, S_{2k})$, 而 $\mathscr{C}_2$ 是至少在两个点上达到极小值的路径.

设 $C_1 \in \mathscr{C}_1$ (图 20), 而 γ 是极小值点. 将路径 $C_1 = (S_0, S_1, \cdots, S_{2k})$ 对应于一正路径 C_1^* (图 21). 路径 C_1^* 是由如下方式得到的: 将轨道 $(S_0, S_1, \cdots, S_l)$ 关于过点

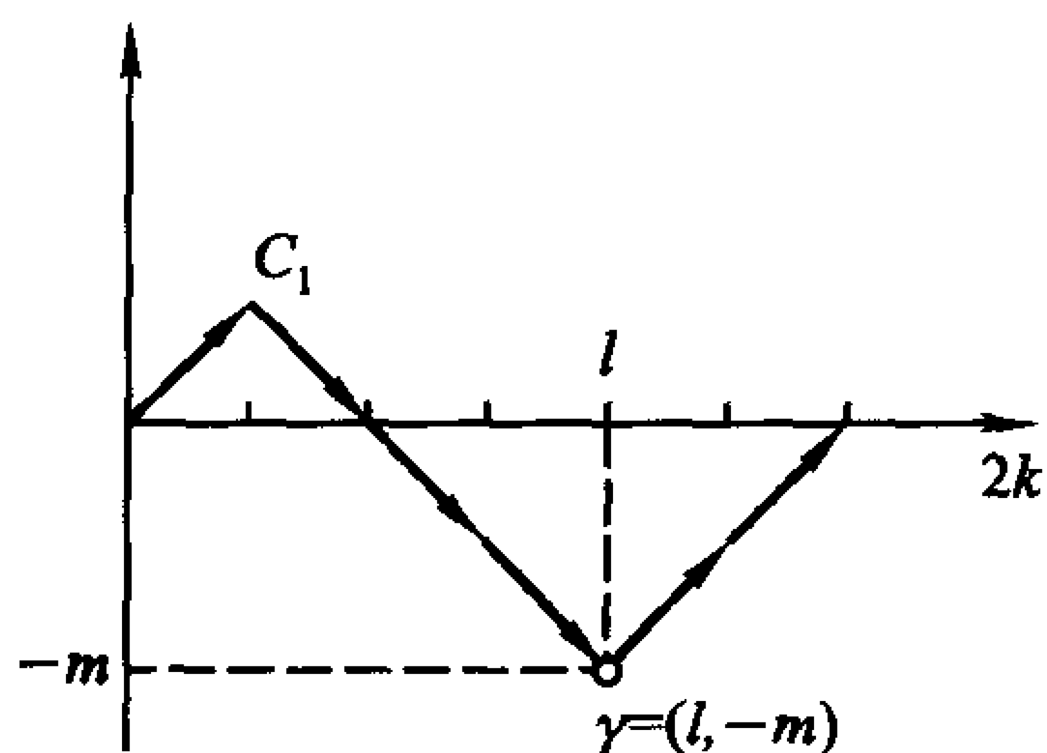


图 20

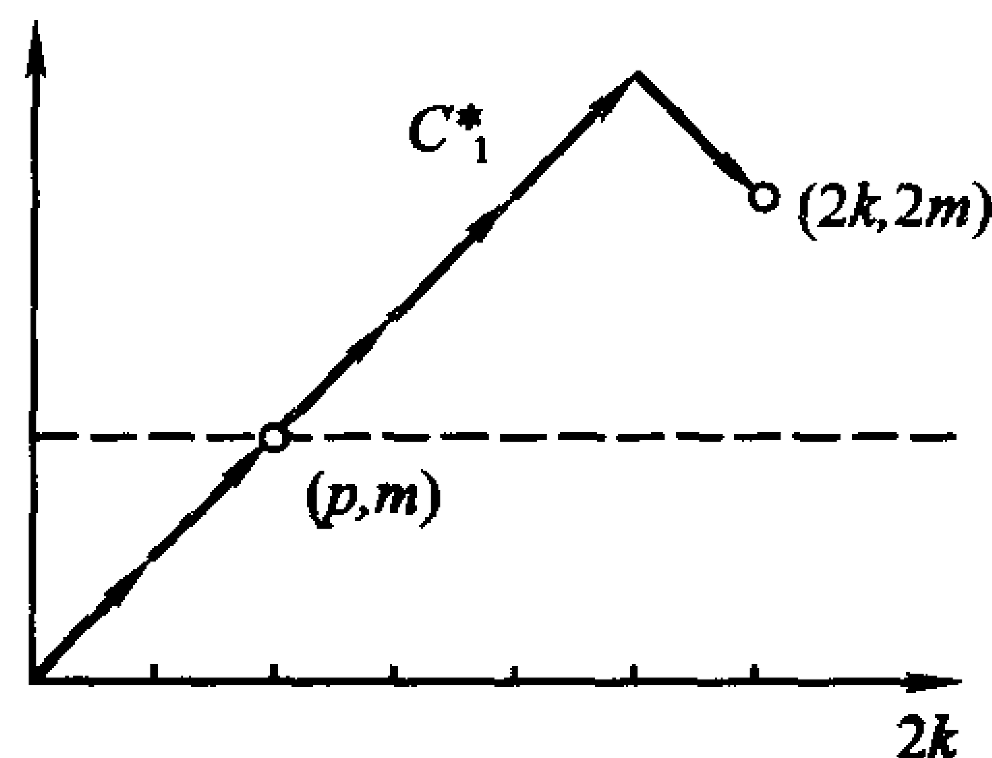


图 21

l 铅直线进行反射, 而把所得轨道自点 $(2k,0)$ 延伸后, 将其置于右、上方. 然后将坐标原点置于点 $(l,-m)$. 所得到的轨道是正路径.

同样, 如果路径 $C_2 \in \mathscr{C}_2$, 则用同样的方式可以将其对应于某一非负路径 C_2^*.

相反, 设 $C_1^* = (S_1 > 0, \cdots, S_{2k} > 0)$ 是某一正路径, 其中 $S_{2k} = 2m$ (图 21). 将路径 C_1^* 对应于一正路径 C_1 (图 21). 路径 C_1 是由如下方式得到的: 设 p 是使 $S_p = m$ 的最后一点; 将轨道 $(S_p, \cdots, S_{2m})$ 关于铅直线 $x = p$ 反射, 将其置于右、上方, 使其第一个端点与点 $(0,0)$ 重合. 然后将坐标原点置于所得轨道的左端点 (这恰好是图 20 上轨道). 所得路径有唯一极小值, 且 $S_{2k} = 0$. 对路径

$$(S_1 \geqslant 0, \cdots, S_{2k} \geqslant 0, \text{ 且 } \exists i, 0 \leqslant i \leqslant 2k, S_i = 0)$$

施用类似的构造可以导致至少有两个极小值且 $S_{2k} = 0$ 的路径. 从而建立了一一对应关系, 于是证明了 (11) 式.

这样, 证明了等式 (9), 以及下面的公式:

$$f_{2k} = u_{2(k-1)} - u_{2k} = \frac{1}{2k} u_{2(k-1)}.$$

由斯特林 (J. Stirling) 公式, 有

$$u_{2k} = \mathrm{C}_{2k}^k \times 2^{-2k} \sim \frac{1}{\sqrt{\pi k}}, \quad k \to \infty.$$

因此

$$f_{2k} \sim \frac{1}{2\sqrt{\pi} k^{3/2}}, \quad k \to \infty.$$

由此可见, 首返 0 的时间的数学期望

$$\mathbf{E} \min(\sigma_{2n}, 2n) = \sum_{k=1}^{n} 2k \mathbf{P}\{\sigma_{2n} = 2k\} + 2n u_{2n} = \sum_{k=1}^{n} u_{2(k-1)} + 2n u_{2n}$$

是相当大的.

此外

$$\sum_{k=1}^{\infty} u_{2(k-1)} = \infty.$$

从而, 对于步数无限的情形, 游动趋于 0 的平均时间的极限等于 ∞.

这一事实可以解释所研究的、对称随机游动许多意想不到的性质. 例如, 两个实力相当的对手 (即 $p=q=1/2$) 对弈时, 自然地期望经 $2n$ 局对弈, 平局的*平均次数* (即使 $S_i=0$ 的时刻 i 的次数) 与 $2n$ 成正比. 然而, 平均平局次数却是数量级为 $\sqrt{2n}$ 的量 (参见第七章 §9, (17) 式). 特别, 由此可以得出与期望相反的、"典型的"的实现: 游动 $(S_0, S_1, \cdots, S_n)$ 事实上并不具有*正弦的*特点 (质点一半时间在正侧, 另一半时间在负侧), 而具有*长而缓慢的波动*的特点. 所谓*反正弦定律*说明了命题的这种提法. 我们现在就开始介绍反正弦定律.

2. 反正弦定律 记 $P_{2k,2n}$ 为质点在时段上 $[0,2n]$ 的*正方*度过 $2k$ 个单位时间的概率*).

引理 2 设 $u_0=1$ 和 $0 \leqslant k \leqslant n$, 则

$$P_{2k,2n}=u_{2k}u_{2n-2k}. \tag{12}$$

证明 上面曾证明 $f_{2k}=u_{2(k-1)}-u_{2k}$. 现在证明,

$$u_{2k}=\sum_{r=1}^{k} f_{2r}u_{2(k-r)}. \tag{13}$$

由于 $\{S_{2k}=0\} \subseteq \{\sigma_{2n} \leqslant 2k\}$, 可见

$$\{S_{2k}=0\}=\{S_{2k}=0\}\cap\{\sigma_{2n}\leqslant 2k\}=\sum_{1\leqslant l\leqslant k}\{S_{2k}=0\}\cap\{\sigma_{2n}=2l\}.$$

从而

$$\begin{aligned}u_{2k}=\mathbf{P}\{S_{2k}=0\}&=\sum_{1\leqslant l\leqslant k}\mathbf{P}\{S_{2k}=0,\sigma_{2n}=2l\}\\&=\sum_{1\leqslant l\leqslant k}\mathbf{P}\{S_{2k}=0|\sigma_{2n}=2l\}\mathbf{P}\{\sigma_{2n}=2l\}.\end{aligned}$$

而由

$$\begin{aligned}\mathbf{P}\{S_{2k}=0|\sigma_{2n}=2l\}&=\mathbf{P}\{S_{2k}=0|S_1\neq 0,\cdots,S_{2l-1}\neq 0,S_{2l}=0\}\\&=\mathbf{P}\{S_{2l}+(\xi_{2l+1}+\cdots+\xi_{2k})=0|S_1\neq 0,\cdots,S_{2l-1}\neq 0,S_{2l}=0\}\\&=\mathbf{P}\{S_{2l}+(\xi_{2l+1}+\cdots+\xi_{2k})=0|S_{2l}=0\}\\&=\mathbf{P}\{\xi_{2l+1}+\cdots+\xi_{2k}=0\}=\mathbf{P}\{S_{2(k-1)}=0\},\end{aligned}$$

可见

$$u_{2k}=\sum_{1\leqslant l\leqslant k}\mathbf{P}\{S_{2(k-l)}=0\}\mathbf{P}\{\sigma_{2n}=2l\},$$

*) 称质点在时间段上 $[m-1,m]$ 位于正方, 如果 S_{m-1} 或 S_m 至少有一个值是正的.

于是,(13) 式得证.

现在证明 (12) 式. 当 $k=0$ 和 $k=n$ 时 (12) 式显然成立. 现在设 $1 \leqslant k \leqslant n-1$. 假如质点在正方度过 $2k$ 个时刻, 则它一定通过 0. 设 $2r$ 是首次返回 0 的时刻. 有两种可能的情形: 当 $S_i \geqslant 0$ 时 $l \leqslant 2r$, 当 $S_i \leqslant 0$ 时 $l \leqslant 2r$.

易见, 对于第一种情形, 路径条数等于

$$\left(\frac{1}{2} \times 2^{2r} f_{2r}\right) 2^{2(n-r)} P_{2(k-r),2(n-r)} = \frac{1}{2} \times 2^{2n} f_{2r} P_{2(k-r),2(n-r)}.$$

对于第二种情形, 相应的路径条数等于

$$\frac{1}{2} \times 2^{2n} f_{2r} P_{2k,2(n-r)}.$$

从而, 对于 $1 \leqslant k \leqslant n-1$, 有

$$P_{2k,2n} = \frac{1}{2} \sum_{r=1}^{k} f_{2r} P_{2(k-r),2(n-r)} + \frac{1}{2} \sum_{r=1}^{k} f_{2r} P_{2k,2(n-r)}. \tag{14}$$

假设对于 $m=1,\cdots,n-1$, 公式 $P_{2k,2m} = u_{2m} u_{2m-2k}$ 成立. 那么, 由 (13) 和 (14) 两式, 可见

$$\begin{aligned} P_{2k,2n} &= \frac{1}{2} u_{2n-2k} \sum_{r=1}^{k} f_{2r} u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{k} f_{2r} u_{2n-2r-2k} \\ &= \frac{1}{2} u_{2n-2k} u_{2k} + \frac{1}{2} u_{2k} u_{2n-2k} = u_{2k} u_{2n-2k}. \end{aligned}$$

□

现在, 设 $\gamma(2n)$ 是质点在区间 $[0,2n]$ 内位于正方的时间单位数, 则对于 $x<1$, 有

$$\mathbf{P}\left\{\frac{1}{2} < \frac{\gamma(2n)}{2n} \leqslant x\right\} = \sum_{\left\{k: \frac{1}{2} < \frac{2k}{2n} \leqslant x\right\}} P_{2k,2n}.$$

由于当 $k \sim \infty$ 时, 有

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}},$$

则当 $k \to \infty, n-k \to \infty$ 时

$$P_{2k,2n} = u_{2k} u_{2(n-k)} \sim \frac{1}{\pi \sqrt{k(n-k)}}.$$

因此

$$\sum_{\left\{k: \frac{1}{2} < \frac{2k}{2n} \leqslant x\right\}} P_{2k,2n} - \sum_{\left\{k: \frac{1}{2} < \frac{2k}{2n} \leqslant x\right\}} \frac{1}{\pi n} \left[\frac{k}{n}\left(1-\frac{k}{n}\right)\right]^{-1/2} \to 0, n \to \infty,$$

故

$$\sum_{\left\{k:\frac{1}{2}<\frac{2k}{2n}\leqslant x\right\}} P_{2k,2n}-\frac{1}{\pi}\int_{1/2}^{x}\frac{\mathrm{d}t}{\sqrt{t(1-t)}}\to 0, n\to\infty.$$

而由对称性, 可见

$$\sum_{\left\{k:\frac{k}{n}\leqslant\frac{1}{2}\right\}} P_{2k,2n}\to\frac{1}{2}, n\to\infty$$

且

$$\frac{1}{\pi}\int_{1/2}^{x}\frac{\mathrm{d}t}{\sqrt{t(1-t)}}=\frac{2}{\pi}\arcsin\sqrt{x}-\frac{1}{2}.$$

于是, 证明了下面的定理.

定理 (反正弦定律) 质点位于正方的时间所占的份额不大于 x 的概率, 趋于 $2\pi^{-1}\arcsin\sqrt{x}$:

$$\sum_{\left\{k:\frac{k}{n}\leqslant x\right\}} P_{2k,2n}\to\frac{2}{\pi}\arcsin\sqrt{x}. \tag{15}$$

注意, 积分中

$$\int_{0}^{x}\frac{\mathrm{d}t}{\sqrt{t(1-t)}}$$

的被积函数 $u(t)=[t(1-t)]^{-1/2}$ 是在点 $t=0$ 和 $t=1$ 趋向无穷的一条 U 形曲线.

由此可见, 对于充分大的 n, 有

$$\mathbf{P}\left\{0<\frac{\gamma(2n)}{2n}\leqslant\Delta\right\}>\mathbf{P}\left\{\frac{1}{2}<\frac{\gamma(2n)}{2n}\leqslant\frac{1}{2}+\Delta\right\},$$

即形象地说, 质点位于正方接近 0 (或 1), 比自然期望的值 1/2, 更有可能.

利用反正弦函数表, 以及在 (15) 式中的收敛速度实际上是很快的情况, 可以求得下面的概率:

$$\mathbf{P}\left\{\frac{\gamma(2n)}{2n}\leqslant 0.024\right\}\approx 0.1,$$

$$\mathbf{P}\left\{\frac{\gamma(2n)}{2n}\leqslant 0.1\right\}\approx 0.2,$$

$$\mathbf{P}\left\{\frac{\gamma(2n)}{2n}\leqslant 0.2\right\}\approx 0.3,$$

$$\mathbf{P}\left\{\frac{\gamma(2n)}{2n}\leqslant 0.65\right\}\approx 0.6.$$

这样, 如果考虑时段 $[0,1000]$, 则大致在十分之一的情形下, 质点总共有 24 个时间单位在正方度过, 但多数时间 (976 时间单位) 在负方.

3. 练习题

1. 当 $n \to \infty$ 时, $\mathbf{E}\min(\sigma_{2n}, 2n) \to \infty$ 的速度如何?

2. 设 $\tau_n = \min\{1 \leqslant k \leqslant n : S_k = 1\}$, 如果对于一切 $1 \leqslant k \leqslant n, S_k < 1$, 设 $\tau_n = \infty$. 问对于对称 $(p = q = 1/2)$ 和非对称 $(p \neq q)$ 游动, 当 $n \to \infty$ 时, $\mathbf{E}\min(\tau_n, n)$ 的极限如何?

3. 基于 §10 的思路和方法, 对于伯努利对称 $(p = q = 1/2)$ 随机游动 $\{S_k, k \leqslant n\}$, 其中 $S_0 = 0, S_k = \xi_1 + \cdots + \xi_k$, 证明满足下列关系式 ($N$ 是正整数):

$$\mathbf{P}\{\max_{1\leqslant k\leqslant n} S_k \geqslant N, S_n < N\} = \mathbf{P}\{S_n > N\},$$
$$\mathbf{P}\{\max_{1\leqslant k\leqslant n} S_k \geqslant N\} = 2\mathbf{P}\{S_n \geqslant N\} - \mathbf{P}\{S_n = N\},$$
$$\mathbf{P}\{\max_{1\leqslant k\leqslant n} S_k = N\} = \mathbf{P}\{S_n = N\} + \mathbf{P}\{S_n = N+1\}.$$

§11. 鞅. 鞅对随机游动的某些应用

1. 引言 前面研究的伯努利变量 $\xi_1, \cdots, \xi_n$, 形成独立随机变量序列. 在这一节和下一节将引进形成鞅和马尔可夫 A. A. Марков 链的两类非独立随机变量.

鞅论将在第七章中详细介绍. 而现在仅给出定义, 证明关于对于停止时间保持鞅性的一个定理, 给出鞅用于证明所谓关于 "表决" 的定理. 同样地, 该定理将用于 §10 中 (5) 式命题的另一个证明, 该命题曾经利用反射定理证明.

2. 鞅的定义和例 设 $(\Omega, \mathscr{A}, \mathbf{P})$ 是有限概率空间, $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \cdots \preccurlyeq \mathscr{D}_n$ 是某一分割序列.

定义 1 随机变量序列 $\xi = (\xi_k)_{1\leqslant k\leqslant n}$ (关于分割 $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \cdots \preccurlyeq \mathscr{D}_n$) 称做**鞅**, 如果:

1) ξ_k 是 $\mathscr{D}_k-$ 可测的,

2) $\mathbf{E}(\xi_{k+1}|\mathscr{D}_k) = \xi_k, 1 \leqslant k \leqslant n-1$.

为强调随机变量 $\xi = (\xi_1, \cdots, \xi_n)$ 的哪个分割系是鞅, 我们还将使用记号

$$\xi = (\xi_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}, \tag{1}$$

不过为简便计常省略下标 $1 \leqslant k \leqslant n$.

如果分割 $\mathscr{D}_k$ 是由随机变量 $\xi_1, \cdots, \xi_n$ 诱导的, 即

$$\mathscr{D}_k = \mathscr{D}_{\xi_1,\cdots,\xi_n},$$

我们往往不提 $\xi = (\xi_k, \mathscr{D}_k)$ 是鞅, 干脆称序列本身 $\xi = (\xi_k)$ 是鞅.

下面举几个鞅的例子.

例 1 设 $\eta_1, \cdots, \eta_n$ 是独立伯努利随机变量:

$$\mathbf{P}\{\eta_k = 1\} = \mathbf{P}\{\eta_k = -1\} = \frac{1}{2},$$
$$S_k = \eta_1 + \cdots + \eta_k, \mathscr{D}_k = \mathscr{D}_{\eta_1, \cdots, \eta_k}.$$

注意, 分割 $\mathscr{D}_k$ 的构造十分简单:

$$\mathscr{D}_1 = \{D^+, D^-\},$$

其中 $D^+ = \{\omega : \eta_1 = +1\}, D^- = \{\omega : \eta_1 = -1\}$;

$$\mathscr{D}_2 = \{D^{++}, D^{+-}, D^{-+}, D^{--}\},$$

其中 $D^{++} = \{\omega : \eta_1 = +1, \eta_2 = +1\}, \cdots, D^{--} = \{\omega : \eta_1 = -1, \eta_2 = -1\}$; 依此类推.

易见, $\mathscr{D}_{\eta_1, \cdots, \eta_k} = \mathscr{D}_{S_1, \cdots, S_k}$.

现在证明, 序列 $(S_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 构成鞅.

随机变量 S_k 是 $\mathscr{D}_k-$ 可测的, 而由 §8 的 (12) 式, (18) 式和 (24) 式, 可见

$$\mathbf{E}(S_{k+1}|\mathscr{D}_k) = \mathbf{E}(S_{k+\eta_{k+1}}|\mathscr{D}_k) = \mathbf{E}(S_k|\mathscr{D}_k) + \mathbf{E}(\eta_{k+1}|\mathscr{D}_k) = S_k + \mathbf{E}\eta_{k+1} = S_k,$$

而这正是鞅的性质所要求的.

假如设 $S_0 = 0$, 取平凡分割: $\mathscr{D}_0 = \{\Omega\}$, 则序列 $(S_k, \mathscr{D}_k)_{0 \leqslant k \leqslant n}$ 仍然是鞅.

例 2 设 $\eta_1, \cdots, \eta_n$ 是独立伯努利随机变量: $\mathbf{P}\{\eta_i = 1\} = p, \mathbf{P}\{\eta_i = -1\} = q$. 如果 $p \neq q$, 则每一个满足

$$\xi_k = \left(\frac{q}{p}\right)^{S_k}, \xi_k = S_k - k(p-q), \quad \text{其中 } S_k = \eta_1 + \cdots + \eta_k$$

的序列 $\xi = (\xi_k)$ 都是鞅.

例 3 设 η 是独立伯努利随机变量, $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \cdots \preccurlyeq \mathscr{D}_n$, 且

$$\xi_k = \mathbf{E}(\eta|\mathscr{D}_k). \tag{2}$$

那么, 序列 $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 是鞅. 事实上, $\mathbf{E}(\eta|\mathscr{D}_k)$ 显然 $\mathscr{D}_k-$ 可测, 且由 §8 的 (20) 式, 可见

$$\mathbf{E}(\xi_{k+1}|\mathscr{D}_k) = \mathbf{E}[\mathbf{E}(\eta|\mathscr{D}_{k+1})|\mathscr{D}_k] = \mathbf{E}(\eta|\mathscr{D}_k) = \xi_k.$$

因此, 我们指出, 如果 $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 是任意鞅, 则由 §8 的 (20) 式, 可见

$$\xi_k = \mathbf{E}(\xi_{k+1}|\mathscr{D}_k) = \mathbf{E}[\mathbf{E}(\xi_{k+2}|\mathscr{D}_{k+1})|\mathscr{D}_k] = \mathbf{E}(\xi_{k+2}|\mathscr{D}_k) = \cdots = \mathbf{E}(\xi_n|\mathscr{D}_k). \tag{3}$$

于是, 一切鞅 $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 的集合, 包含所有形如 (2) 式的鞅. (注意, 对于无穷序列 $\xi = (\xi_k, \mathscr{D}_k)_{k \geqslant 1}$, 一般并非如此; 参见第七章 §1 练习题 6.)

例 4　设 $\eta_1, \cdots, \eta_n$ 是独立同分布机变量序列, $S_k = \eta_1 + \cdots + \eta_k$,

$$\mathscr{D}_1 = \mathscr{D}_{S_n}, \mathscr{D}_2 = \mathscr{D}_{S_n, S_{n-1}}, \cdots, \mathscr{D}_n = \mathscr{D}_{S_n, \cdots, S_1}.$$

证明序列 $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$, 其中

$$\xi_1 = \frac{S_n}{n}, \xi_2 = \frac{S_{n-1}}{n-1}, \cdots, \xi_k = \frac{S_{n+1-k}}{n+1-k}, \cdots, \xi_n = S_1,$$

是鞅. 首先, 显然 $\mathscr{D}_k \preccurlyeq \mathscr{D}_{k+1}$ 且 ξ_k 为 $\mathscr{D}_k-$ 可测; 其次, 由于对称性, 对于 $j \leqslant n-k+1$, 有

$$\mathbf{E}(\eta_j | \mathscr{D}_k) = \mathbf{E}(\eta_1 | \mathscr{D}_k) \tag{4}$$

(与 §8 中 (26) 式比较). 因此

$$(n-k+1)\mathbf{E}(\eta_1 | \mathscr{D}_k) = \sum_{j=1}^{n-k+1} \mathbf{E}(\eta_j | \mathscr{D}_k) = \mathbf{E}(S_{n-k+1} | \mathscr{D}_k) = S_{n-k+1}.$$

于是

$$\xi_k = \frac{S_{n-k+1}}{n-k+1} = \mathbf{E}(\eta_1 | \mathscr{D}_k),$$

故由例 3 可见序列 $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 是鞅.

注　由已证明的序列 $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 的鞅性可知, 为何有时称 $(S_k/k)_{1 \leqslant k \leqslant n}$ 构成逆鞅或"反向鞅"的原因 (对照第七章 §1 练习题 5).

例 5　设 $\eta_1, \cdots, \eta_n$ 是独立伯努利随机变量:

$$\mathbf{P}\{\eta_i = +1\} = \mathbf{P}\{\eta_i = -1\} = \frac{1}{2}, S_k = \eta_1 + \cdots + \eta_k.$$

设 A 和 B 是两个整数, $A < 0 < B$. 那么, 对于任意 $0 < \lambda < \pi/2$, 序列 $\xi = (\xi_k, \mathscr{D}_k)$, 其中 $\mathscr{D}_k = \mathscr{D}_{S_1, \cdots, S_k}$, 且

$$\xi_k = (\cos \lambda)^{-k} \exp\left\{ i\lambda \left(S_k - \frac{A+B}{2} \right) \right\} \tag{5}$$

构成复鞅 (ξ_k 的实部和虚部都是鞅).

3. 停止时间　由鞅的定义, 可见对于一切 k, 数学期望 $\mathbf{E}\xi_k$ 相同:

$$\mathbf{E}\xi_k = \mathbf{E}\xi_1.$$

事实上, 如果将 (确定性) 时间 k 换成所谓停止时间, 鞅的这一性质仍然成立. 对于相应的提法, 我们引进下面的定义.

定义 2　随机变量 $\tau = \tau(\omega)$ (关于分割 $(\mathscr{D}_k)_{1 \leqslant k \leqslant n}, \mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \cdots \preccurlyeq \mathscr{D}_n$) 称做停止时间, 如果对于 $k = 1, \cdots, n$, 随机变量 $I_{\{\tau = k\}}(\omega)$ 是 $\mathscr{D}_k-$ 可测的.

假如把分割 $\mathscr{D}_k$ 视为 k 步观测结果诱导的 (例如, 由随机变量 $\eta_1, \cdots, \eta_k$ 诱导的分割 $\mathscr{D}_k = \mathscr{D}_{\eta_1, \cdots, \eta_k}$), 则变量 $I_{\{\tau=k\}}(\omega)$ 的 $\mathscr{D}_k-$ 可测性表示: 事件 $\{\tau = k\}$ 存在与否仅取决于 k 步的观测结果 (而不依赖于 "将来").

如果 $\mathscr{B}_k = \alpha(\mathscr{D}_k)$, 则变量 $I_{\{\tau=k\}}(\omega)$ 的 $\mathscr{D}_k-$ 可测性等价于假设:

$$\{\tau = k\} \in \mathscr{B}_k. \tag{6}$$

我们已经遇到过停止时间的例子: §9 和 §10 引进的时间 τ_k^x, σ_{2n}, 就是停止时间的例. 上面列举的停止时间是形如

$$\begin{aligned} \tau^A &= \min\{0 < k \leqslant n : \xi_k \in A\}, \\ \sigma^A &= \min\{0 \leqslant k \leqslant n : \xi_k \in A\} \end{aligned} \tag{7}$$

的停止时间的特殊情形, (相应为在 0 之后首次到达 0 和首次) 到达某一序列 ξ_0, $\xi_1, \cdots, \xi_n$ 的集合 A 的时间.

4. 停止时间的应用

定理 1 设 $\xi = (\xi_k, \mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 是鞅, 而 τ 是关于分割 $(\mathscr{D}_k)_{1 \leqslant k \leqslant n}$ 的某一停止时间, 则

$$\mathbf{E}(\xi_\tau | \mathscr{D}_1) = \xi_1, \tag{8}$$

其中

$$\xi_\tau = \sum_{k=1}^{n} \xi_k I_{\{\tau=k\}} \tag{9}$$

而

$$\mathbf{E}\xi_\tau = \mathbf{E}\xi_1. \tag{10}$$

证明 (仿照 §9 中 (29) 式的证明). 设 $D \in \mathscr{D}_1$, 则利用条件数学期望的性质, 并注意到 (3) 的鞅性, 有

$$\begin{aligned} \mathbf{E}(\xi_\tau | D) &= \frac{\mathbf{E}(\xi_\tau I_D)}{\mathbf{P}(D)} = \frac{1}{\mathbf{P}(D)} \sum_{l=1}^{n} \mathbf{E}(\xi_l I_{\{\tau=l\}} I_D) \\ &= \frac{1}{\mathbf{P}(D)} \sum_{l=1}^{n} \mathbf{E}[\mathbf{E}(\xi_n | \mathscr{D}_l) I_{\{\tau=l\}} I_D] = \frac{1}{\mathbf{P}(D)} \sum_{l=1}^{n} \mathbf{E}[\mathbf{E}(\xi_n I_{\{\tau=l\}} I_D | \mathscr{D}_l)] \\ &= \frac{1}{\mathbf{P}(D)} \sum_{l=1}^{n} \mathbf{E}[\xi_n I_{\{\tau=l\}} I_D] = \frac{1}{\mathbf{P}(D)} \mathbf{E}(\xi_n I_D) = \mathbf{E}(\xi_n | D), \end{aligned}$$

从而

$$\mathbf{E}(\xi_\tau | \mathscr{D}_1) = \mathbf{E}(\xi_n | \mathscr{D}_1) = \xi_1.$$

由此显然地得等式 $\mathbf{E}\xi_\tau = \mathbf{E}\xi_1$. □

系 对于例 1 的鞅 $(S_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$ 和任意 (关于 $\mathscr{D}_k$ 的) 停止时间 τ, 有等式

$$\mathbf{E}S_\tau = 0, \quad \mathbf{E}S_\tau^2 = \mathbf{E}\tau. \tag{11}$$

等式 (11) 称做瓦尔德 (A. Wald) 恒等式 (对照 §9, (29) 式和 (30) 式; 以及第七章 §2 的练习题 1 和定理 3).

利用定理 1 证明如下命题.

定理 2 ("表决" 定理) 设 $\eta_1, \cdots, \eta_n$ 是独立同分布机变量序列, 其中 η_i 在集合上 $\{0, 1, \cdots\}$ 只取有限个值, $S_k = \eta_1 + \cdots + \eta_k, 1 \leqslant k \leqslant n$, 则

$$\mathbf{P}\{S_k < k, k \in [1, n] | S_n\} = \left(1 - \frac{S_n}{n}\right)^+, \tag{12}$$

其中 $a^+ = \max(a, 0)$.

证明 (12) 式在集合 $\{\omega : S_n \geqslant n\}$ 上显然. 因此, 只需证明对于使 $S_n < n$ 的基本事件证明 (12).

考虑在例 4 中引进的鞅 $\xi = (\xi_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$, 其中

$$\xi_k = \frac{S_{n-k+1}}{n-k+1}, \quad \mathscr{D}_k = \mathscr{D}_{S_{n+1-k}, \cdots, S_n}.$$

记

$$\tau = \min\{1 \leqslant k \leqslant n : \xi_k \geqslant 1\},$$

在集合

$$\{\xi_k < 1, k \in [1, n]\} = \left\{\max_{1\leqslant l\leqslant n} \frac{S_l}{l} < 1\right\}$$

上设 $\tau = n$. 显然, 在此集合上 $\xi_\tau = \xi_n = S_1 = 0$, 因而

$$\left\{\max_{1\leqslant l\leqslant n} \frac{S_l}{l} < 1\right\} = \left\{\max_{1\leqslant l\leqslant n} \frac{S_l}{l} < 1, S_n < n\right\} \subseteq \{\xi_\tau = 0\}. \tag{13}$$

现在考虑同时使

$$S_n < n, \max_{1\leqslant l\leqslant n} \frac{S_l}{l} \geqslant 1$$

的基本事件. 记 $\sigma = n + 1 - \tau$. 易见,

$$\sigma = \min\{1 \leqslant k \leqslant n : S_k \geqslant k\},$$

因此 (由于 $S_n < n$) $\sigma < n, S_\sigma \geqslant \sigma$, 而 $S_{\sigma+1} < \sigma + 1$. 可见 $\eta_{\sigma+1} = S_{\sigma+1} - S_\sigma < (\sigma + 1) - \sigma = 1$, 即 $\eta_{\sigma+1} = 0$. 所以 $\sigma \leqslant S_\sigma = S_{\sigma+1} < \sigma + 1$, 因而 $S_\sigma = \sigma$ 且

$$\xi_\tau = \frac{S_{n+1-\tau}}{n+1-\tau} = \frac{S_\sigma}{\sigma} = 1.$$

从而

$$\left\{\max_{1\leqslant l\leqslant n}\frac{S_l}{l}\geqslant 1, S_n<n\right\}\subseteq\{\xi_\tau=1\}. \tag{14}$$

由 (13) 式和 (14) 式, 可见

$$\left\{\max_{1\leqslant l\leqslant n}\frac{S_l}{l}\geqslant 1, S_n<n\right\}=\{\xi_\tau=1\}\cap\{S_n<n\}.$$

因此, 在集合 $S_n<n$ 上

$$\mathbf{P}\left\{\max_{1\leqslant l\leqslant n}\frac{S_l}{l}\geqslant 1\,\middle|\, S_n\right\}=\mathbf{P}\{\xi_\tau=1|S_n\}=\mathbf{E}\{\xi_\tau|S_n\},$$

其中因为 ξ_τ 只有 0 和 1 两个可能值, 所以最后一个等式成立.

最后, 注意到, $\mathbf{E}(\xi_\tau|S_n)=\mathbf{E}(\xi_\tau|\mathscr{D}_1)$, 且由定理 1 知 $\mathbf{E}(\xi_\tau|\mathscr{D}_1)=\xi_1=S_n/n$. 于是, 在集合 $\{S_n<n\}$ 上, 有

$$\mathbf{P}\{S_k<k, k\in[1,n]|S_n\}=1-\frac{S_n}{n}.$$ □

我们用定理 2 给出 §10 中引理 1 的另一证明, 并说明表决定理名称的来历.

设 $\xi_1,\cdots,\xi_n$ 是独立伯努利随机变量序列, 且

$$\mathbf{P}\{\xi_i=1\}=\mathbf{P}\{\xi_i=-1\}=\frac{1}{2},$$

$S_k=\xi_1+\cdots+\xi_k$, 而 a 和 b 非负正数: $a-b>0, a+b=n$. 现在证明,

$$\mathbf{P}\{S_1>0,\cdots,S_n>0|S_n=a-b\}=\frac{a-b}{a+b}. \tag{15}$$

事实上, 由对称性, 可见

$$\begin{aligned}
&\mathbf{P}\{S_1>0,\cdots,S_n>0|S_n=a-b\}\\
&=\mathbf{P}\{S_1<0,\cdots,S_n<0|S_n=-(a-b)\}\\
&=\mathbf{P}\{S_1+1<1,\cdots,S_n+n<n|S_n+n=n-(a-b)\}\\
&=\mathbf{P}\{\eta_1<1,\cdots,\eta_1+\cdots+\eta_n<n|\eta_1+\cdots+\eta_n=n-(a-b)\}\\
&=\left[1-\frac{n-(a-b)}{n}\right]^+=\frac{a-b}{n}=\frac{a-b}{a+b},
\end{aligned}$$

其中设 $\eta_k=\xi_k+1$, 并用到了 (12) 式.

§10 中的 (5) 式曾经借助反射原理证明的, 现在显然可以由式 (15) 导出 §10 中的 (5) 式.

假设有两个候选人 A 和 B, 我们可以把 $\xi_i=1$ 视为投 A 一票, 而把 $\xi_i=-1$ 视为投 B 一票. 假如有 k 个人参加投票, 而 S_k 是 A 和 B 得票的差额; 假设 A 最后共得 a 票, 而 B 最后共得 b 票, 且 $a-b>0, a+b=n$, 则

$$\mathbf{P}\{S_1>0,\cdots,S_n>0|S_n=a-b\}$$

表示 A 得票始终领先于 B 的概率. 根据 (15) 此概率等于 $(a-b)/n$.

5. 练习题

1. 设, $\mathscr{D}_0 \preccurlyeq \mathscr{D}_1 \preccurlyeq \cdots \preccurlyeq \mathscr{D}_n$ 是分割序列, $\mathscr{D}_0 = \{\Omega\}; \eta_k$ 是 $\mathscr{D}_k-$ 可测随机变量, $1 \leqslant k \leqslant n$. 证明序列 $\xi = (\xi_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$, 其中

$$\xi_k = \sum_{l=1}^{k} [\eta_l - \mathbf{E}(\eta_l | \mathscr{D}_{l-1})]$$

是鞅.

2. 设对于随机变量 $\eta_1, \cdots, \eta_k, \mathbf{E}(\eta_k | \eta_1, \cdots, \eta_{k-1}) = 0$. 证明序列 $\xi = (\xi_k)_{1\leqslant k\leqslant n}$, 其中 $\xi_1 = \eta_1$ 且

$$\xi_{k+1} = \sum_{i=1}^{k} \eta_{i+1} f_i(\eta_1, \cdots, \eta_i),$$

其中 f_i 是构成鞅的某一函数.

3. 证明, 任何鞅 $\xi = (\xi_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$ 都有不相关增量: 对于 $a < b < c < d$, 有

$$\operatorname{cov}(\xi_d - \xi_c, \xi_b - \xi_a) = 0.$$

4. 设对于随机变量序列 $\xi = (\xi_1, \cdots, \xi_n)$, 其中 ξ_k 为 $\mathscr{D}_k-$ 可测 $(\mathscr{D}_1 \preccurlyeq \mathscr{D}_2 \preccurlyeq \cdots \preccurlyeq \mathscr{D}_n)$. 证明, 为使序列 $\xi = (\xi_1, \cdots, \xi_n)$ (关于分割系 $\{\mathscr{D}_k\}$) 为鞅的充分和必要条件是, 对于任意 (关于 $\{\mathscr{D}_k\}$ 的) 停止时间 τ, 有 $\mathbf{E}\xi_\tau = \mathbf{E}\xi_1$. (条件 "对于任意停止时间", 可以换成条件 "对于只有两个可能值的停止时间".)

5. 证明, 如果 $\xi = (\xi_k, \mathscr{D}_k)_{1\leqslant k\leqslant n}$ 是鞅, 而 τ 是停止时间, 则对于任意 k,

$$\mathbf{E}[\xi_n I_{\{\tau=k\}}] = \mathbf{E}[\xi_k I_{\{\tau=k\}}].$$

6. 设 $\xi = (\xi_k, \mathscr{D}_k)$ 和 $\eta = (\eta_k, \mathscr{D}_k)$ 是两个鞅, 且 $\xi_1 = \eta_1 = 0$. 证明

$$\mathbf{E}\xi_n\eta_n = \sum_{k=2}^{n} \mathbf{E}(\xi_k - \xi_{k-1})(\eta_k - \eta_{k-1}),$$

特别,

$$\mathbf{E}\xi_n^2 = \sum_{k=2}^{n} \mathbf{E}(\xi_k - \xi_{k-1})^2.$$

7. 设 $\eta_1, \cdots, \eta_n$ 是独立同分布机变量序列, 其中 $\mathbf{E}\eta_i = 0$. 证明序列 $\xi = (\xi_k)$, 其中

$$\xi_k = \left(\sum_{i=1}^{k} \eta_i\right)^2 - k\mathbf{E}\eta_1^2,$$

$$\xi_k = \frac{\exp\{\lambda(\eta_1 + \cdots + \eta_k)\}}{[\mathbf{E}\exp\{\lambda\eta_1\}]^k}$$

都是鞅.

8. 设 $\eta_1, \cdots, \eta_n$ 是独立同分布随机变量序列, 其值域为有限集合 Y. 设

$$f_0(y) = \mathbf{P}\{\eta_1 = y\} > 0, y \in Y; f_1(y) \geqslant 0, \sum_{y \in Y} f_1(y) = 1.$$

证明

$$\xi = (\xi_k, \mathscr{D}_k^\eta), \quad \xi_k = \frac{f_1(\eta_1)\cdots f_1(\eta_k)}{f_0(\eta_1)\cdots f_0(\eta_k)}, \quad \mathscr{D}_k^\eta = \mathscr{D}_{\eta_1,\cdots,\eta_k}$$

是鞅. (关系式 ξ_k 称做似然比, 在数理统计中有十分重要的应用.)

§12. 马尔可夫链. 遍历性定理. 强马尔可夫性

1. 马尔可夫链 在以上研究的伯努利概型中, $\Omega = \{\omega : \omega = (x_1, \cdots, x_n), x_i = 0, 1\}$, 每个基本事件 ω 的概率为式 $\mathbf{P}(\{\omega\}) = p(\omega)$, 其中

$$p(\omega) = p(x_1)\cdots p(x_n), \quad p(x) = p^x q^{1-x}. \tag{1}$$

在此情况下随机变量 $\xi_1, \cdots, \xi_n$ 独立同分布, 其中 $\xi_i(\omega) = x_i$, 且

$$\mathbf{P}\{\xi_1 = x\} = \cdots = \mathbf{P}\{\xi_n = x\} = p(x), \quad x = 0, 1.$$

如果将 (1) 式换成

$$p(\omega) = p_1(x_1)\cdots p_n(x_n),$$

其中 $p_i(x) = p_i^x(1-p_i)^{1-x}, 0 \leqslant p_i \leqslant 1$, 则这时随机变量 $\xi_1, \cdots, \xi_n$ 不再是独立的, 且一般也不是同分布的:

$$\mathbf{P}\{\xi_1 = x\} = p_1(x), \cdots, \mathbf{P}\{\xi_n = x\} = p_n(x).$$

我们考虑这类概型的一种推广, 将引出由非独立随机变量形成的所谓的马尔可夫链.

假设

$$\Omega = \{\omega : \omega = (x_0, x_1, \cdots, x_n), x_i \in X\},$$

其中 X 是某一有穷集合. 假设给定非负函数: $p_0(x), p_1(x, y), \cdots, p_n(x, y)$, 满足

$$\begin{aligned} &\sum_{x \in X} p_0(x) = 1, \\ &\sum_{y \in X} p_k(x, y) = 1, \quad k = 1, \cdots, n;\ x \in X. \end{aligned} \tag{2}$$

对于每一 $\omega = (x_0, x_1, \cdots, x_n)$, 设 $\mathbf{P}(\{\omega\}) = p(\omega)$, 其中

$$p(\omega) = p_0(x_0)p_1(x_0, x_1)\cdots p_n(x_{n-1}, x_n). \tag{3}$$

不难验证

$$\sum_{\omega \in \Omega} p(\omega) = 1,$$

全体 $p(\omega)$, 以及空间 Ω 连同其一切子集的集系决定空间 $(\Omega, \mathscr{A}, \mathbf{P})$, 通常称之为形成**马尔可夫链的试验模型**.

引进随机变量 $\xi_0, \xi_1, \cdots, \xi_n$, 其中对于 $\omega = (x_1, \cdots, x_n), \xi_i(\omega) = x_i$. 经简单计算, 可得

$$\begin{gathered}\mathbf{P}\{\xi_0 = a\} = p_0(a), \\ \mathbf{P}\{\xi_0 = a_0, \xi_1 = a_1, \cdots, \xi_k = a_k\} = p_0(a_0)p_1(a_0, a_1)\cdots p_k(a_{k-1}, a_k).\end{gathered} \tag{4}$$

现在对于上述概率模型 $(\Omega, \mathscr{A}, \mathbf{P})$, 证明条件概率的一条重要性质: 假设 $\mathbf{P}\{\xi_k = a_k, \cdots, \xi_0 = a_0\} > 0$,

$$\mathbf{P}\{\xi_{k+1} = a_{k+1} | \xi_k = a_k, \cdots, \xi_0 = a_0\} = \mathbf{P}\{\xi_{k+1} = a_{k+1} | \xi_k = a_k\}. \tag{5}$$

由于 (4) 式和条件概率的定义 (§3), 可见

$$\begin{aligned}\mathbf{P}\{\xi_{k+1} = a_{k+1} | \xi_k = a_k, \cdots, \xi_0 = a_0\} &= \frac{\mathbf{P}\{\xi_{k+1} = a_{k+1}, \cdots, \xi_0 = a_0\}}{\mathbf{P}\{\xi_k = a_k, \cdots, \xi_0 = a_0\}} \\ &= \frac{p_0(a_0)p_1(a_0, a_1)\cdots p_{k+1}(a_k, a_{k+1})}{p_0(a_0)p_1(a_0, a_1)\cdots p_k(a_{k-1}, a_k)} = p_{k+1}(a_k, a_{k+1}).\end{aligned}$$

类似可得

$$\mathbf{P}\{\xi_{k+1} = a_{k+1} | \xi_k = a_k\} = p_{k+1}(a_k, a_{k+1}), \tag{6}$$

因而性质 (5) 得证.

设 $\mathscr{D}_k^\xi = \mathscr{D}_{\xi_0, \cdots, \xi_k}$ 是随机变量 $\xi_0, \cdots, \xi_k$ 诱导的分割, 而 $\mathscr{B}_k^\xi = \alpha(\mathscr{D}_k^\xi)$.

那么, 根据 §8 引进的记号, 由 (5) 式可见

$$\mathbf{P}\{\xi_{k+1} = a_{k+1} | \mathscr{B}_k^\xi\} = \mathbf{P}\{\xi_{k+1} = a_{k+1} | \xi_k\} \tag{7}$$

或

$$\mathbf{P}\{\xi_{k+1} = a_{k+1} | \xi_0, \cdots, \xi_k\} = \mathbf{P}\{\xi_{k+1} = a_{k+1} | \xi_k\}.$$

注 鉴于 (5) 式和 (7) 式以及 0 **概率事件**, 暂时中断我们的叙述, 作对整个以后的全部内容至关重要的说明.

在推导 (5) 式时曾经假设 $\mathbf{P}\{\xi_k = a_k, \cdots, \xi_0 = a_0\} > 0$ (即 $\mathbf{P}\{\xi_k = a_k\} > 0$). 之所以需要这样, 是因为条件概率 $\mathbf{P}(A|B)$ (暂时!) 只有在 $\mathbf{P}(B) > 0$ 的条件下才有定义.

但是, 需要指出, 如果 $B = \{\xi_k = a_k, \cdots, \xi_0 = a_0\}$ 而且 $\mathbf{P}(B) = 0$ (因此, 对于 $C = \{\xi_k = a_k\}, \mathbf{P}(C) = 0$), 则 "路径" 应视为**不可实现的**, 那么, 事件 $\{\xi_{k+1} = a_k\}$ 关于这条**不可实现的** "路径" 的条件概率的问题, 也就毫无实际意义.

因此, 为明确起见, 我们以后把条件概率定义为:

$$\mathbf{P}(A|B) = \begin{cases} \dfrac{\mathbf{P}(AB)}{\mathbf{P}(B)}, & \text{若 } \mathbf{P}(B) > 0, \\ 0, & \text{若 } \mathbf{P}(B) = 0. \end{cases}$$

在这样的定义下, 无需诸如 $\mathbf{P}\{\xi_k = a_k, \cdots, \xi_0 = a_0\} > 0$ 的任何前提, (5) 式和 (7) 式仍然是正确的.

应强调指出, 所指出的与零概率事件有关的困难, 在概率论的论述中相当典型. 在第二章 §7 中将引进 (关于任意分割、$\sigma-$ 代数 $\cdots\cdots$ 的) 条件概率的一般定义, 可以自然地 "用于零概率的" 情形.

如果利用明显的等式

$$\mathbf{P}(AB|C) = \mathbf{P}(A|BC)\mathbf{P}(B|C),$$

则由 (7) 式, 有

$$\mathbf{P}\{\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1}|\mathscr{B}_k^\xi\} = \mathbf{P}\{\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1}|\xi_k\}, \tag{8}$$

或

$$\mathbf{P}\{\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1}|\xi_0, \cdots, \xi_k\} = \mathbf{P}\{\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1}|\xi_k\}. \tag{9}$$

此等式有如下直观的解释. 假设 ξ_k 表示质点在 "现在", $(\xi_0, \cdots, \xi_{k-1})$ 表示质点在 "过去", $(\xi_{k+1}, \cdots, \xi_n)$ 表示质点在 "将来". 那么, (9) 式表示: 在 "过去" $(\xi_0, \cdots, \xi_{k-1})$ 和 "现在" ξ_k 固定的情况下, "将来" $(\xi_{k+1}, \cdots, \xi_n)$ 仅依赖于 "现在" ξ_k, 而与质点如何到达点 ξ_k 无关, 即 "将来" 不依赖于 "过去".

设 $\mathrm{J} = \{\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1}\}, \mathrm{X} = \{\xi_k = a_k\}, \mathrm{G} = \{\xi_{k-1} = a_{k-1}, \cdots, \xi_0 = a_0\}$①, 则由 (9) 式可见

$$\mathbf{P}(\mathrm{J}|\mathrm{XG}) = \mathbf{P}(\mathrm{J}|\mathrm{X}\},$$

由此容易求出

$$\mathbf{P}(\mathrm{JG}|\mathrm{X}) = \mathbf{P}(\mathrm{J}|\mathrm{X}\}\mathbf{P}(\mathrm{G}|\mathrm{X}). \tag{10}$$

换句话说, 由 (7) 式可见, 在固定 "现在" X 的情况下, "将来" J 和 "过去" G **相互独立**. 不难证明逆命题: 如果对于任意 $k = 0, 1, \cdots, n-1$, (10) 式成立, 则对于任意 $k = 0, 1, \cdots, n-1$, 性质 (7) 也成立.

"将来" 和 "过去" 的独立性, 或者说, 在固定 "现在" 的情况下, "将来" 和 "过去" 相互独立, 通常称做**马尔可夫性**, 而相应的随机变量序列 $\xi_1, \cdots, \xi_n$ 称做**马尔可夫链**.

① J 表示 "将来", X 表示 "现在", G 表示 "过去". —— 译者

这样, 如果由 (3) 式决定基本事件的 “权重” $p(\omega)$, 则序列 $\xi=(\xi_0,\cdots,\xi_n)$ 构成马尔可夫链, 其中 $\xi_i(\omega)=x_i$.

于是, 引出如下定义.

定义 设 $(\Omega,\mathscr{A},\mathbf{P})$ 是某一 (有限) 概率空间, $\xi=(\xi_0,\cdots,\xi_n)$ 是 (有限) 随机变量序列, 而 (有限) 集合 X 是其值域. 如果满足条件 (7), 则序列 $\xi=(\xi_0,\cdots,\xi_n)$ 称做 (有限) *马尔可夫链*.

集合 X 称做马尔可夫链的*相空间*或*状态空间*. 概率的全体

$$\{p_0(x)\},p_0(x)=\mathbf{P}\{\xi_0=x\},x\in X$$

称为*初始分布*; 而矩阵

$$(p_k(x,y)),x,y\in X,p_k(x,y)=\mathbf{P}\{\xi_k=y|\xi_{k-1}=x\}$$

称为在时刻 $k=1,\cdots,n$(自状态 x 到状态 y) 的*转移概率矩阵*.

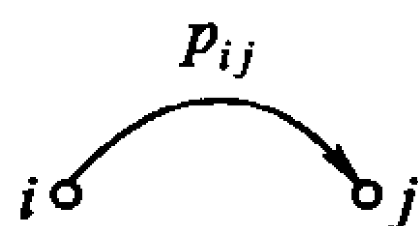
如果转移概率*不依赖于* $k: p_k(x,y)=p(x,y)$, 则序列 $\xi=(\xi_0,\cdots,\xi_n)$ 称为转移概率矩阵是 $(p(x,y))$ 的*齐次*马尔可夫链.

注意, $(p(x,y))$ 是随机矩阵: 其元素非负, 任何一行元素之和等于 1,

$$\sum_y p(x,y)=1,\quad x\in X.$$

假设相空间 X 是有限个整数点的集合 (例如, $X=\{0,1,\cdots,N\},X=\{0,\pm1,\cdots,\pm N\}$ 等). 此外, 根据习惯记 $p_i=p_0(i),p_{ij}=p(i,j)$.

显然, 齐次马尔可夫链的性质, 完全决定于初始分布 p_i 和转移概率 p_{ij}. 在有些情形下, 为描绘马尔可夫链的演变, 不明显写出矩阵 (p_{ij}), 而是运用有向*网络*, 其结点表示 X 中的状态.



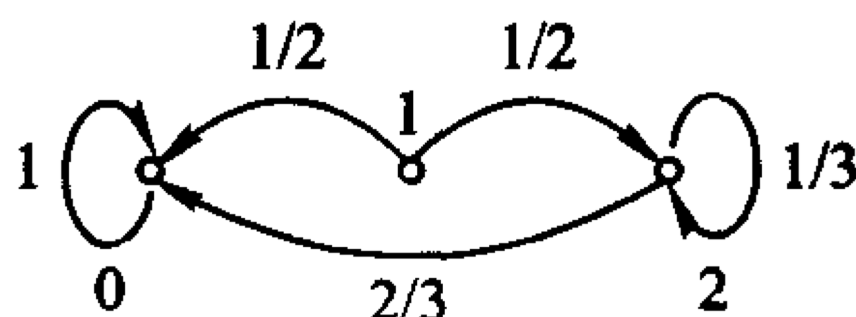
而箭头表示自状态 i 指向状态 j, 箭线上的 p_{ij} 是质点可能自点 i 转移到点 j 的概率.

例 1 设 $X=\{0,1,2\}$ 和

$$(p_{ij})=\begin{pmatrix}1&0&0\\1/2&0&1/2\\2/3&0&1/3\end{pmatrix}.$$

该矩阵对应下面的网络图:

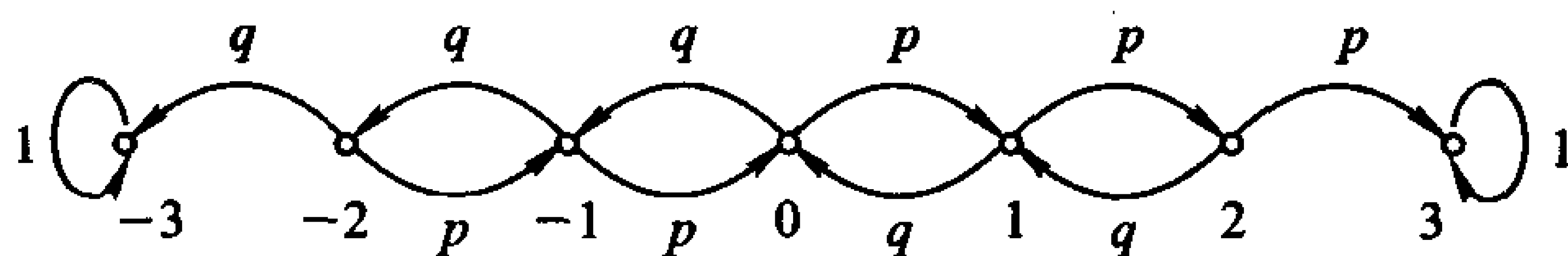
注意, 这里状态 0 是 “吸收状态”: 由于 $p_{00}=1$, 可见一旦质点进入状态 0, 则它就停留在此状态. 质点自状态 1 以相同的概率进入相邻的状态 0 或 2; 对于状态 2, 质点留在其中的概率为 1/3, 而由状态 2 转移到状态 0 的概率为 2/3.



例 2 设 $X=\{0,\pm1,\cdots,\pm N\}, p_0=1, p_{NN}=p_{-N,-N}=1$, 而对于 $|i|<N$,

$$p_{ij}=\begin{cases}p, & 若\ j=i+1,\\ q, & 若\ j=i-1,\\ 0, & 其他情形.\end{cases} \tag{11}$$

可以用图形表示对应于这样链的转移情况 $(N=3)$:



该链对应于上面研究的甲、乙两人的博弈: 每人的赌金为 N, 甲每一步以概率 p 赢乙得 $+1$, 以概率 q 输给乙得 -1. 假如把状态 i 视为甲赢得乙的金额, 则到达状态 N 和 $-N$, 分别表示乙破产和甲破产.

事实上, 如果 $\eta_1,\eta_2,\cdots,\eta_n$ 是相互独立的伯努利随机变量, 其中 $\mathbf{P}\{\eta_i=+1\}=p, \mathbf{P}\{\eta_i=-1\}=q$, 而 $S_k=\eta_1+\eta_2+\cdots+\eta_k$ 是甲赢得乙的金额, 则序列 $S_0,S_1,\cdots,S_n$ $(S_0=0)$ 形成马尔可夫链且 $p_0=1$, 而转移概率矩阵为 (11), 因为

$$\begin{aligned}&\mathbf{P}\{S_{k+1}=j|S_k=i_k,S_{k-1}=i_{k-1},\cdots,S_1=i_1\}\\ =&\mathbf{P}\{S_k+\eta_{k+1}=j|S_k=i_k,S_{k-1}=i_{k-1},\cdots,S_1=i_1\}\\ =&\mathbf{P}\{S_k+\eta_{k+1}=j|S_k=i_k\}=\mathbf{P}\{\eta_{k+1}=j-i_k\}.\end{aligned}$$

马尔可夫链 $S_0,S_1,\cdots,S_n$ 具有十分简单的构造:

$$S_{k+1}=S_k+\eta_{k+1},\quad 0\leqslant k\leqslant n-1,$$

其中 $\eta_1,\eta_2,\cdots,\eta_n$ 是独立随机变量序列.

同理可得, 如果 $\xi_0,\eta_1,\cdots,\eta_n$ 是独立随机变量序列, 则序列 $\xi_0,\xi_1,\cdots,\xi_n$ 也是马尔可夫链, 其中

$$\xi_{k+1}=f_k(\xi_k,\eta_{k+1}),\quad 0\leqslant k\leqslant n-1. \tag{12}$$

因此应该指出, 自然把这样构造的马尔可夫链, 视为由递推关系式

$$x_{k+1}=f_k(x_k)$$

决定的, 如 (确定性) 序列 $(x_0,x_1,\cdots,x_n)$ 的概率的类似.

现在再举一个 "排队论" 中、形如 (12) 式的马尔可夫链的例子.

例 3 假设在出租车站每单位时间有一辆车到达. 假如在停车站无乘客等候, 则车立即离去. 以 η_k 表示于 k 时到达车站等车的乘客数量, 并且假定 $\eta_1, \eta_2, \cdots, \eta_n$ 是相互独立的随机变量. 设 ξ_k 为在时刻 k 乘客的队长, 其中 $\xi_0 = 0$. 那么, 如果 $\xi_k = i$, 则在下一时刻 $k+1$ 的队长等于

$$j = \begin{cases} \eta_{k+1}, & 若\ i = 0, \\ i - 1 + \eta_{k+1}, & 若\ i \geqslant 1. \end{cases}$$

换句话说,

$$\xi_{k+1} = (\xi_k - 1)^+ + \eta_{k+1}, \quad 0 \leqslant k \leqslant n-1,$$

其中 $a^+ = \max(a, 0)$, 从而序列 $\xi = (\xi_0, \cdots, \xi_n)$ 是马尔可夫链.

例 4 此例涉及分支过程理论. 所谓离散时间分支过程, 是这样一个随机变量序列 $\xi_0, \xi_1, \cdots, \xi_n$, 其中 ξ_k 表示在时刻 k 质点的数量, 而且质点的生与灭的情况是按如下方式进行的: 每个质点以概率 $p_j (j = 0, 1, \cdots, M)$ 裂变为 j 个质点, 裂变既不依赖于其他质点, 也与 "过去历史" 无关.

假设在初始时刻总共只有一个质点, $\xi_0 = 1$. 假如在时刻 k 有编号为 $1, 2, \cdots, \xi_k$ 的 ξ_k 个质点, 那么根据上面的描述, ξ_{k+1} 等于随机多个随机变量之和:

$$\xi_{k+1} = \eta_1^{(k)} + \cdots + \eta_{\xi_k}^{(k)},$$

其中 $\eta_i^{(k)}$ 是第 i 号质点裂变的质点的个数. 显然, 若 $\xi_0 = 0$, 则 $\xi_{k+1} = 0$. 假设所有随机变量 $\eta_j^{(k)}, k \geqslant 0, j \geqslant 1$ 相互独立, 则

$$\begin{aligned} &\mathbf{P}\{\xi_{k+1} = i_{k+1} | \xi_k = i_k, \xi_{k-1} = i_{k-1}, \cdots\} \\ &= \mathbf{P}\{\xi_{k+1} = i_{k+1} | \xi_k = i_k\} = \mathbf{P}\{\eta_1^{(k)} + \cdots + \eta_{i_k}^{(k)} = i_{k+1}\}. \end{aligned}$$

由此可见, 序列 $\xi_0, \xi_1, \cdots, \xi_n$ 是马尔可夫链.

特别重要的是下面的情形: 每一个质点, 或者是以概率 q 灭绝, 或者以概率 p 裂变为两个质点, 其中 $p + q = 1$. 对于这种情形, 容易经计算得到: 转移概率

$$p_{ij} = \mathbf{P}\{\xi_{k+1} = j | \xi_k = i\}$$

由下面的公式给出.

$$p_{ij} = \begin{cases} \mathrm{C}_i^{j/2} p^{j/2} q^{i-j/2}, & 若 j = 0, \cdots, 2i, \\ 0, & 其他. \end{cases}$$

2. 柯尔莫戈洛夫 – 查普曼 (Kolmogorov-Chapman) 方程 记 $\xi = (\xi_k, \Pi, \mathbb{P})$ 为齐次马尔可夫链, 其初始概率行向量为 $\Pi = (p_i)$, 而转移概率矩阵为 $\mathbb{P} = (p_{ij})$. 显然.

$$p_{ij} = \mathbf{P}\{\xi_1 = j | \xi_0 = i\} = \cdots = \mathbf{P}\{\xi_n = j | \xi_{n-1} = i\}.$$

记

$$p_{ij}^{(k)} = \mathbf{P}\{\xi_k = j|\xi_0 = i\} \quad (= \mathbf{P}\{\xi_{k+l} = j|\xi_l = i\}, l = 1, 2, \cdots)$$

为经 k 步由状态 i 到状态 j 的转移概率, 而

$$p_j^{(k)} = \mathbf{P}\{\xi_k = j\}$$

是质点在 k 时位于 j 点的概率. 亦设

$$\Pi^{(k)} = (p_i^{(k)}), \quad \mathbb{P}^{(k)} = (p_{ij}^{(k)}).$$

我们证明, 转移概率满足 "*柯尔莫戈洛夫 – 查普曼方程*":

$$p_{ij}^{(k+l)} = \sum_\alpha p_{i\alpha}^{(k)} p_{\alpha j}^{(l)}, \tag{13}$$

其矩阵形式为

$$\mathbb{P}^{(k+l)} = \mathbb{P}^{(k)}\mathbb{P}^{(l)}, \tag{14}$$

基于全概率公式和马尔可夫性, 关系式 (13) 很容易证明:

$$\begin{aligned} p_{ij}^{(k+l)} &= \mathbf{P}\{\xi_{k+l} = j|\xi_0 = i\} = \sum_\alpha \mathbf{P}\{\xi_{k+l} = j, \xi_k = \alpha|\xi_0 = i\} \\ &= \sum_\alpha \mathbf{P}\{\xi_{k+l} = j|\xi_k = \alpha\}\mathbf{P}\{\xi_k = \alpha|\xi_0 = i\} = \sum_\alpha p_{\alpha j}^{(l)} p_{i\alpha}^{(k)}. \end{aligned}$$

方程 (13) 的两种特殊情形最重要: *后向方程*

$$p_{ij}^{(l+1)} = \sum_\alpha p_{i\alpha} p_{\alpha j}^{(l)} \tag{15}$$

和*前向方程*

$$p_{ij}^{(k+1)} = \sum_\alpha p_{i\alpha}^{(k)} p_{\alpha j} \tag{16}$$

(见图 22 和图 23).

前向方程和后向方程的矩阵形式相应为:

$$\mathbb{P}^{(k+1)} = \mathbb{P}^{(k)}\mathbb{P}, \tag{17}$$

$$\mathbb{P}^{(k+1)} = \mathbb{P}\mathbb{P}^{(k)}. \tag{18}$$

对于 (无条件) 概率 $p_j^{(k)}$, 类似地可以得到

$$p_j^{(k+l)} = \sum_\alpha p_\alpha^{(k)} p_{\alpha j}^{(l)}, \tag{19}$$

而其矩阵形式为

$$\Pi^{(k+l)} = \Pi^{(k)}\mathbb{P}^{(l)}.$$

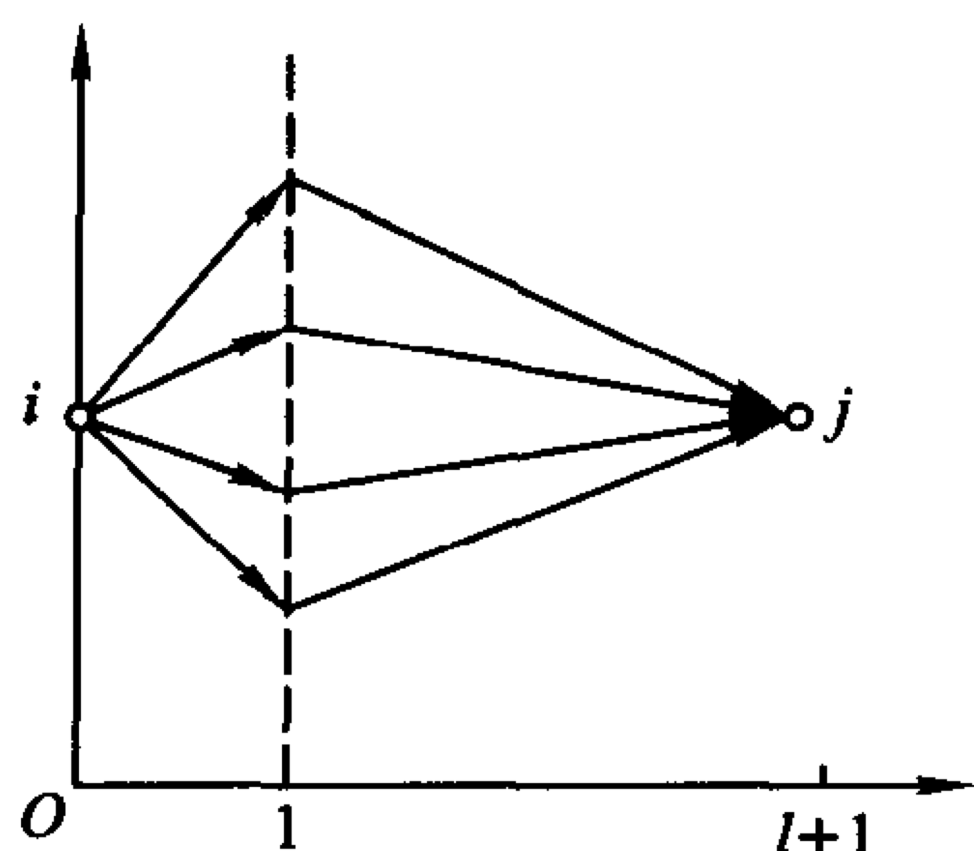


图 22 后向方程

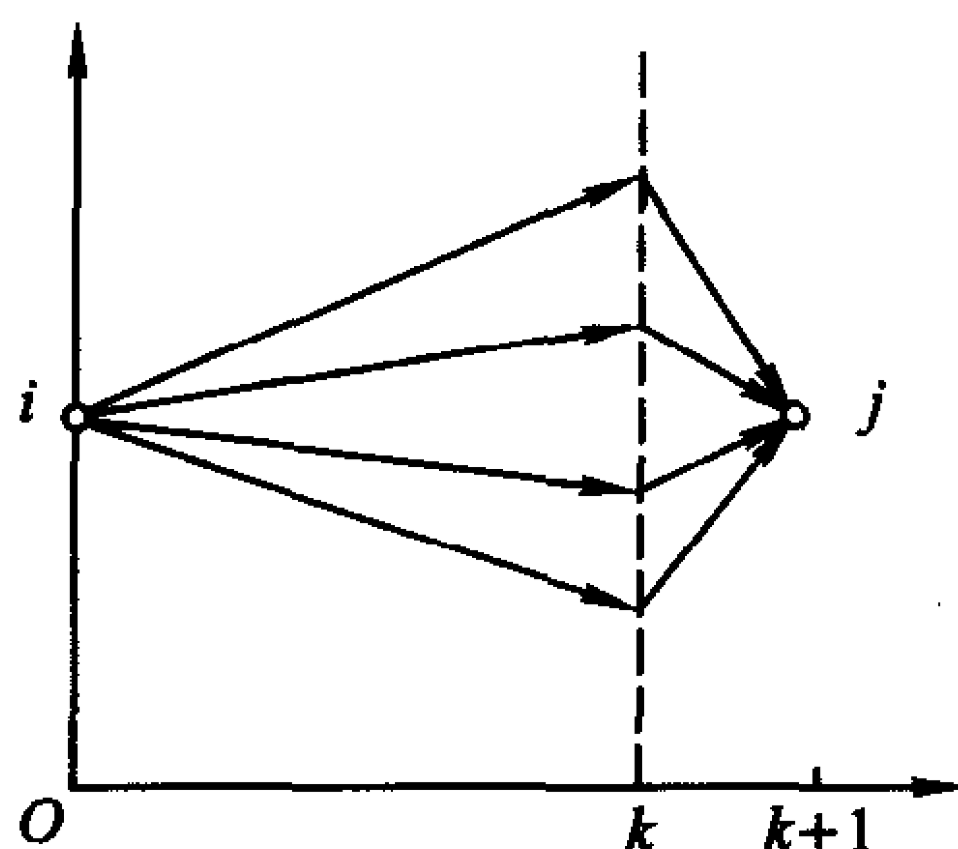


图 23 前向方程

特别, 前向方程为

$$\Pi^{(k+1)} = \Pi^{(k)}\mathbb{P},$$

而后向方程为

$$\Pi^{(k+1)} = \Pi^{(1)}\mathbb{P}^{(k)}.$$

由于 $\mathbb{P}^{(1)} = \mathbb{P}, \Pi^{(0)} = \Pi$, 可见

$$\mathbb{P}^{(k)} = \mathbb{P}^k, \quad \Pi^{(k)} = \Pi\mathbb{P}^k.$$

因此, 对于齐次马尔可夫链, k 步的转移概率 $p_{ij}^{(k)}$ 是矩阵 $\mathbb{P}$ 的 k 次幂. 从而, 齐次马尔可夫链的许多性质, 可以用矩阵方法进行研究.

例 考虑具有 0 和 1 两个状态, 而转移矩阵为

$$\mathbb{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$$

的齐次马尔可夫链. 不难计算

$$\mathbb{P}^2 = \begin{pmatrix} p_{00}^2 + p_{01}p_{10} & p_{01}(p_{00} + p_{11}) \\ p_{10}(p_{00} + p_{11}) & p_{11}^2 + p_{01}p_{10} \end{pmatrix},$$

而由归纳法, 有

$$\mathbb{P}^n = \frac{1}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{11} & 1 - p_{00} \\ 1 - p_{11} & 1 - p_{00} \end{pmatrix} + \frac{(p_{00} + p_{11} - 1)^n}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{00} & -(1 - p_{00}) \\ -(1 - p_{11}) & 1 - p_{11} \end{pmatrix}$$

(假设 $|p_{00} + p_{11} - 1| < 1$).

由此可见, 如果矩阵 $\mathbb{P}$ 的元素使 $|p_{00}+p_{11}-1|<1$ (特别, 如所有转移概率 p_{ij} 都大于 0), 则当 $n\to\infty$ 时, 有

$$\mathbb{P}^n \to \frac{1}{2-p_{00}-p_{11}}\begin{pmatrix}1-p_{11} & 1-p_{00}\\ 1-p_{11} & 1-p_{00}\end{pmatrix}, \tag{20}$$

从而

$$\lim_n p_{i0}^{(n)}=\frac{1-p_{11}}{2-p_{00}-p_{11}},\quad \lim_n p_{i1}^{(n)}=\frac{1-p_{00}}{2-p_{00}-p_{11}}.$$

于是, 如果 $|p_{00}+p_{11}-1|<1$, 则所研究的马尔可夫链的行为有如下规律性: 随着时间的推移, 初始状态对 "质点处于不同状态的概率" 的影响逐渐消失 ($p_{ij}^{(n)}$ 收敛于不依赖于 i 的极限值 π_j, 并形成概率分布: $\pi_0\geqslant 0,\pi_1\geqslant 0,\pi_0+\pi_1=1$); 假如除此之外假设 $p_{ij}>0$, 则极限值 $\pi_0>0,\pi_1>0$ (对照下面的定理 1).

3. 遍历性 下面的定理描绘的广泛的一类马尔可夫链, 具有所谓*遍历性*: 极限

$$\pi_j=\lim_n p_{ij}^{(n)}$$

不但存在且不依赖于 i 构成概率分布

$$\pi_j>0,\quad \sum_j \pi_j=1,$$

而且对于所有 $j,\pi_j>0$ (这样的分布称做*遍历分布*, 详见第八章 §3).

定理 1 (遍历性定理) *设 $\mathbb{P}=(p_{ij})$ 是马尔可夫链的转移概率矩阵, 且状态 $X=\{1,2,\cdots,N\}$ 有穷.*

a) 如果对于某个 n_0, 有

$$\min_{i,j} p_{ij}^{(n_0)}>0, \tag{21}$$

则存在 $\pi_1,\cdots,\pi_N$, 使

$$\pi_j>0,\sum_j \pi_j=1, \tag{22}$$

而且对每个 $j\in X$ 和任意 $i\in X$, 有

$$p_{ij}^{(n)}\to\pi_j, n\to\infty. \tag{23}$$

b) 相反, 如果存在满足条件 (22) 和 (23) 的数 $\pi_1,\cdots,\pi_N$, 则存在满足条件 (21) 的 n_0.

c) 式 (22) 中的数 $(\pi_1,\cdots,\pi_N)$, 满足方程组

$$\pi_j=\sum_j \pi_\alpha p_{\alpha j},\quad j=1,\cdots,N. \tag{24}$$

证明　a) 记

$$m_j^{(n)} = \min_i p_{ij}^{(n)}, \quad M_j^{(n)} = \max_i p_{ij}^{(n)}.$$

由于

$$p_{ij}^{(n+1)} = \sum_\alpha p_{i\alpha} p_{\alpha j}^{(n)}, \tag{25}$$

可见

$$m_j^{(n+1)} = \min_i p_{ij}^{(n+1)} = \min_i \sum_\alpha p_{i\alpha} p_{\alpha j}^{(n)} \geqslant \min_i \sum_\alpha p_{i\alpha} \min_\alpha p_{\alpha j}^{(n)} = m_j^{(n)},$$

因此 $m_j^{(n)} \leqslant m_j^{(n+1)}$, 且类似地 $M_j^{(n)} \geqslant M_j^{(n+1)}$. 从而, 为证明命题 (23), 只需证明:

$$M_j^{(n)} - m_j^{(n)} \to 0, n \to \infty, j = 1, \cdots, N.$$

设 $\varepsilon = \min_{i,j} p_{ij}^{(n_0)} > 0$, 则

$$\begin{aligned}
p_{ij}^{(n_0+n)} &= \sum_\alpha p_{i\alpha}^{(n_0)} p_{\alpha j}^{(n)} = \sum_\alpha [p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}] p_{\alpha j}^{(n)} + \varepsilon \sum_\alpha p_{j\alpha}^{(n)} p_{\alpha j}^{(n)} \\
&= \sum_\alpha [p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}] p_{\alpha j}^{(n)} + \varepsilon p_{jj}^{(2n)},
\end{aligned}$$

而由于 $p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)} \geqslant 0$, 可见

$$p_{ij}^{(n_0+n)} \geqslant m_j^{(n)} \sum_\alpha [p_{i\alpha}^{(n_0)} - \varepsilon p_{j\alpha}^{(n)}] + \varepsilon p_{jj}^{(2n)} = m_j^{(n)}(1-\varepsilon) + \varepsilon p_{jj}^{(2n)},$$

于是

$$m_j^{(n_0+n)} \geqslant m_j^{(n)}(1-\varepsilon) + \varepsilon p_{jj}^{(2n)}.$$

类似可得

$$M_j^{(n_0+n)} \leqslant M_j^{(n)}(1-\varepsilon) + \varepsilon p_{jj}^{(2n)}.$$

由上面两个不等式, 得

$$M_j^{(n_0+n)} - m_j^{(n_0+n)} \leqslant (M_j^{(n)} - m_j^{(n)})(1-\varepsilon),$$

从而

$$M_j^{(kn_0+n)} - m_j^{(kn_0+n)} \leqslant (M_j^{(n)} - m_j^{(n)})(1-\varepsilon)^k \downarrow 0, \quad k \to \infty.$$

于是, 当 $n \to \infty$ 时, 对于某序列 $\{n_\beta\}$, $M_j^{(n_\beta)} - m_j^{(n_\beta)} \to 0$. 由于差 $M_j^{(n)} - m_j^{(n)}$ 关于 n 的单调性可见, 当 $n \to \infty$ 时, $M_j^{(n)} - m_j^{(n)} \to 0$.

若记 $\pi_j = \lim_n m_j^{(n)}$, 则由所得到的估计可见, 对于 $n \geqslant n_0$, 有

$$|p_{ij}^{(n)} - \pi_j| \leqslant M_j^{(n)} - m_j^{(n)} \leqslant (1-\varepsilon)^{[n/n_0]-1},$$

即 $p_{ij}^{(n)}$ 以几何速度收敛于极限值 π_j.

同样显然, $m_j^{(n)} \geqslant m_j^{(n_0)} \geqslant \varepsilon > 0, n \geqslant n_0$, 即 $\pi_j > 0$.

b) 因为状态的个数有限, 且 $\pi_j > 0$, 故由 (23) 式可以直接得到 (21) 式.

c) 由 (23) 式和 (25) 式, 得方程 (24). □

4. 马尔可夫链的大数定律 在马尔可夫链的理论中, 方程组

$$x_j = \sum_j x_\alpha p_{\alpha j}, \quad j = 1, \cdots, N \tag{24*}$$

有重要作用 (对照 (24) 式). 其任何非负解 $\mathbb{Q} = (q_1, \cdots, q_N)$ 都满足条件 $\sum_\alpha q_\alpha = 1$. 习惯上称做转移矩阵为 (p_{ij}) 的, 马尔可夫链的*平稳分布*或*不变分布*. 现在对这一名称作如下说明.

设初始分布为 $\mathbb{Q} = (q_1, \cdots, q_N)$, 即假设 $p_j = q_j, j = 1, \cdots, N$. 那么,

$$p_j^{(1)} = \sum_\alpha q_\alpha p_{\alpha j} = q_j$$

且一般 $p_j^{(n)} = q_j$. 换句话说, 假如取 $\mathbb{Q} = (q_1, \cdots, q_N)$ 做初始分布, 则此分布随时间的改变*无变化*, 即对于任意 k,

$$\mathbf{P}\{\xi_k = j\} = \mathbf{P}\{\xi_0 = j\}, \quad j = 1, \cdots, N.$$

不但如此, 具有这样初始分布 $\mathbb{Q} = (q_1, \cdots, q_N)$ 的马尔可夫链 $\xi = (\xi_k, \mathbb{Q}, \mathbb{P})$ 是*平稳的*: 对于任意 l, 随机变量 $(\xi_k, \xi_{k+1}, \cdots, \xi_{k+l})$ 的联合概率分布不依赖于 k (假设 $k + l \leqslant n$).

条件 (21) 保证了存在不依赖于 i 的极限

$$\pi_j = \lim_n p_{ij}^{(n)},$$

而且遍历分布存在, 即存在 $(\pi_1, \cdots, \pi_N), \pi_j > 0$. 这时分布 $(\pi_1, \cdots, \pi_N)$ 也是平稳分布. 现在证明分布 $(\pi_1, \cdots, \pi_N)$ 是唯一平稳分布.

事实上, 假设 $(\widetilde{\pi}_1, \cdots, \widetilde{\pi}_N)$ 是另一个平稳分布. 那么

$$\widetilde{\pi}_j = \sum_\alpha \widetilde{\pi}_\alpha p_{\alpha j} = \cdots = \sum_\alpha \widetilde{\pi}_\alpha p_{\alpha j}^{(n)}.$$

由于 $p_{\alpha j}^{(n)} \to \pi_j$, 可见

$$\widetilde{\pi}_j = \sum_\alpha (\widetilde{\pi}_\alpha \pi_j) = \pi_j.$$

注意, *对于非遍历链也可能存在平稳概率分布* (并且唯一). 事实上, 若

$$\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

则

$$\mathbb{P}^{2n}=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \mathbb{P}^{2n+1}=\begin{pmatrix}0&1\\1&0\end{pmatrix},$$

从而极限 $\lim_n p_{ij}^{(n)}$ 不存在. 然而, 方程组

$$q_j=\sum_\alpha q_\alpha p_{\alpha j},\quad j=1,2$$

变为方程组

$$\begin{cases}q_1=q_2,\\ q_2=q_1,\end{cases}$$

$(1/2,1/2)$ 是其满足条件 $q_1+q_2=1$ 的唯一解 (q_1,q_2).

还应指出, 对于前面在第 2 小节讨论的例子, 方程组 (24*) 为 (其中 $x_j=q_j$):

$$\begin{cases}q_0=q_0p_{00}+q_1p_{10},\\ q_1=q_0p_{01}+q_1p_{11}.\end{cases}$$

所以, 注意到条件 $q_1+q_2=1$, 由此可见, 唯一平稳分布 (q_1,q_2) 就是已求得的分布:

$$q_0=\frac{1-p_{11}}{2-p_{00}-p_{11}},\quad q_1=\frac{1-p_{00}}{2-p_{00}-p_{11}}.$$

现在, 讨论由遍历性定理得出的若干推论.

设 A 是一状态组, $A\subseteq X$, 而

$$I_A(x)=\begin{cases}1, & 若\ x\in A,\\ 0, & 若\ x\notin A.\end{cases}$$

引进随机变量

$$\nu_A(n)=\frac{I_A(\xi_0)+\cdots+I_A(\xi_n)}{n+1},$$

即质点在集合 A 中度过的时间的比率. 由于

$$\mathbf{E}[I_A(\xi_k)|\xi_0=i]=\mathbf{P}[\xi_k\in A|\xi_0=i]=\sum_{j\in A}p_{ij}^{(k)}\quad(=p_i^{(k)}(A)),$$

可见

$$\mathbf{E}[\nu_A(n)|\xi_0=i]=\frac{1}{n+1}\sum_{k=0}^n p_i^{(k)}(A),$$

特别

$$\mathbf{E}[\nu_{\{j\}}(n)|\xi_0=i]=\frac{1}{n+1}\sum_{k=0}^n p_{ij}^{(k)}.$$

由数学分析 (亦见第四章 §3 的引理 1) 知, 如果序列 $a_n \to a$, 则

$$\frac{a_0+\cdots+a_n}{n+1} \to a, \quad n\to\infty.$$

因此, 如果当 $k\to\infty$ 时 $p_{ij}^{(k)} \to \pi_j$, 则

$$\mathbf{E}\nu_{\{j\}}(n) \to \pi_j, \mathbf{E}\nu_A(n) \to \pi_A, \quad \text{其中 } \pi_A = \sum_{j\in A}\pi_j.$$

实际上, 对于遍历链可以证明得更多, 即可以证明随机变量 $I_A(\xi_0),\cdots,$ $I_A(\xi_n),\cdots$ 服从大数定律.

大数定律. *如果 $\xi_0,\xi_1,\cdots$ 是有限遍历马尔可夫链, 则对于任意 $\varepsilon>0$, 任何 $A\subseteq X$, 以及任意初始分布,*

$$\mathbf{P}\{|\nu_A(n)-\pi_A|>\varepsilon\} \to 0, \quad n\to\infty. \tag{26}$$

在证明之前需要指出, 不能直接利用 §5 中关于伯努利随机变量序列 $I_A(\xi_0),\cdots,$ $I_A(\xi_n),\cdots$ 的结果, 因为一般说来这些变量是不独立的. 然而, 如果仍然利用切比雪夫不等式, 以及如下事实: 对于有限状态遍历链, 存在 $0<\rho<1$ 和 $C>0$, 使

$$|p_{ij}^{(n)}-\pi_j| \leqslant C\rho^n, \tag{27}$$

则可以用与独立变量同样的方法证明 (26) 式.

考虑状态 i 和 j (有可能重合), 并证明对于 $\varepsilon>0$, 有

$$\mathbf{P}\{|\nu_{\{j\}}(n)-\pi_j|>\varepsilon|\xi_0=i\} \to 0, \quad n\to\infty. \tag{28}$$

由切比雪夫不等式, 有

$$\mathbf{P}\{|\nu_{\{j\}}(n)-\pi_j|>\varepsilon|\xi_0=i\} \leqslant \frac{\mathbf{E}\{|\nu_{\{j\}}(n)-\pi_j|^2|\xi_0=i\}}{\varepsilon^2}.$$

因此, 只需证明

$$\mathbf{E}\{|\nu_{\{j\}}(n)-\pi_j|^2|\xi_0=i\} \to 0, \quad n\to\infty.$$

通过简单的计算, 可得

$$\begin{aligned}
&\mathbf{E}\{|\nu_{\{j\}}(n)-\pi_j|^2|\xi_0=i\}\\
&=\frac{1}{(n+1)^2}\mathbf{E}\left\{\left[\sum_{k=0}^n \left(I_{\{j\}}(\xi_k)-\pi_j\right)\right]^2\middle|\xi_0=i\right\} = \frac{1}{(n+1)^2}\sum_{k=0}^n\sum_{l=0}^n m_{ij}^{(k,l)},
\end{aligned}$$

其中

$$\begin{aligned}
m_{ij}^{(k,l)} &= \mathbf{E}\{[I_{\{j\}}(\xi_k)I_{\{j\}}(\xi_l)]|\xi_0=i\} - \pi_j\mathbf{E}[I_{\{j\}}(\xi_k)|\xi_0=i]\\
&\quad -\pi_j\mathbf{E}[I_{\{j\}}(\xi_l)|\xi_0=i] + \pi_j^2\\
&= p_{ij}^{(s)}p_{jj}^{(t)} - \pi_j p_{ij}^{(k)} - \pi_j p_{ij}^{(l)} + \pi_j^2,\\
&\qquad s=\min(k,l), t=|k-l|.
\end{aligned}$$

由 (27) 式

$$p_{ij}^{(n)} = \pi_j + \varepsilon_{ij}^{(n)}, \quad |\varepsilon_{ij}^{(n)}| \leqslant C\rho^n.$$

所以

$$|m_{ij}^{(k,l)}| \leqslant C_1[\rho^s + \rho^t + \rho^k + \rho^l],$$

其中 C_1 是某一常数. 从而

$$\begin{aligned}\frac{1}{(n+1)^2}\sum_{k=0}^{n}\sum_{l=0}^{n} m_{ij}^{(k,l)} &\leqslant \frac{C_1}{(n+1)^2}\sum_{k=0}^{n}\sum_{l=0}^{n}[\rho^s + \rho^t + \rho^k + \rho^l] \\ &\leqslant \frac{4C_1}{(n+1)^2} \times \frac{2(n+1)}{1-\rho} = \frac{8C_1}{(n+1)(1-\rho)} \to 0, \quad n \to \infty.\end{aligned}$$

由此就可以证明关系式 (28), 而由 (28) 式可以明显地得到命题 (26).

5. 游动时间的概率和期望的递推方程 对于由伯努利概型诱导的随机游动 $S_0, S_1, \cdots$, 在 §9 对于越出某一边界时间的概率和数学期望, 曾经得到递推方程. 现在对于马尔可夫链导出类似的方程.

假设 $\xi = (\xi_0, \xi_1, \cdots)$ 是转移概率矩阵为 (p_{ij})、相空间为 $X = \{0, \pm1, \cdots, \pm N\}$ 的马尔可夫链. 设 A 和 B 是两个整数, $-N \leqslant A \leqslant 0 \leqslant B \leqslant N$ 且 $x \in X$. 以 $\mathscr{B}_{k+1}$ 表示经右端点首次越出区间 (A, B) 的轨道 $(x_0, x_1, \cdots, x_k)$, $x_i \in X$ 的集合, 即越出集合 (A, B) 进入集合 $(B, B+1, \cdots, N)$ 的轨道 $(x_0, x_1, \cdots, x_k)$, $x_i \in X$ 的全体.

对于 $A \leqslant x \leqslant B$, 设

$$\beta_k(x) = \mathbf{P}\{(\xi_0, \cdots, \xi_k) \in \mathscr{B}_{k+1} | \xi_0 = x\}.$$

为求这些概率 (即马尔可夫链经右端点首次越出区间 (A, B) 的概率), 可以利用推导后向方程的方法. 有

$$\begin{aligned}\beta_k(x) &= \mathbf{P}\{(\xi_0, \cdots, \xi_k) \in \mathscr{B}_{k+1} | \xi_0 = x\} \\ &= \sum_y p_{xy}\mathbf{P}\{(\xi_0, \cdots, \xi_k) \in \mathscr{B}_{k+1} | \xi_0 = x, \xi_1 = y\},\end{aligned}$$

其中, 利用马尔可夫性和链的齐性, 不难证明

$$\begin{aligned}&\mathbf{P}\{(\xi_0, \cdots, \xi_k) \in \mathscr{B}_{k+1} | \xi_0 = x, \xi_1 = y\} \\ &= \mathbf{P}\{(x, y, \xi_2, \cdots, \xi_k) \in \mathscr{B}_{k+1} | \xi_0 = x, \xi_1 = y\} \\ &= \mathbf{P}\{(y, \xi_2, \cdots, \xi_k) \in \mathscr{B}_k | \xi_1 = y\} \\ &= \mathbf{P}\{(y, \xi_1, \cdots, \xi_{k-1}) \in \mathscr{B} | \xi_0 = y\} = \beta_{k-1}(y).\end{aligned}$$

因此, 对于 $A < x < B, 1 \leqslant k \leqslant n$,

$$\beta_k(x) = \sum_y p_{xy}\beta_{k-1}(y).$$

这时, 显然

$$\beta_k(x)=1,\quad x=B,B+1,\cdots,N,$$

和

$$\beta_k(x)=0,\quad x=-N,\cdots,A.$$

对于经左端点首次越出区间 (A,B) 的概率 $\alpha_k(x)$, 可以类似地导出方程.

设 $\tau_k=\min\{0\leqslant l\leqslant k:\xi_l\notin(A,B)\}$, 并且当集合 $\{\cdot\}=\varnothing$ 时 $\tau_k=k$. 那么, 用对 $m_k(x)=\mathbf{E}\{\tau_k|\xi_0=x\}$ 用的同样方法, 可以导出如下方程:

$$m_k(x)=1+\sum_y m_{k-1}(y)p_{xy}$$

(其中 $1\leqslant k\leqslant n, A<x<B$). 这时

$$m_k(x)=0,\quad x\notin(A,B).$$

显然, 如果转移概率矩阵由 (11) 式给出, 则 $\alpha_k(x),\beta_k(x)$ 和 $m_k(x)$ 的方程就是 §9 中相应的方程, 这些方程的推导方法实质上与这里方法一样.

最重要的是, 在游动时间无限的情况下, 将这些方程用于极限情形. 像 §9 中一样, 当 $k\to\infty$ 时, 相应的方程通过形式地极限过度, 可以由以上得到的方程导出.

作为例子, 考虑状态为 $\{0,1,\cdots,B\}$ 的马尔可夫链, 假设其转移概率为

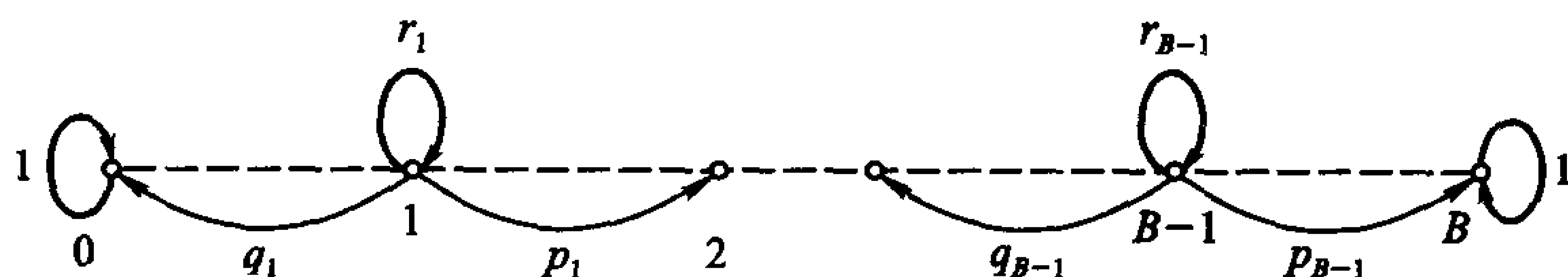
$$p_{00}=1,\quad p_{BB}=1,$$

而对于 $1\leqslant i\leqslant B-1$

$$p_{ij}=\begin{cases}p_i>0, & 若\ j=i+1,\\ r_i, & 若\ j=i,\\ q_i>0, & 若\ j=i-1,\end{cases}$$

其中 $p_i+r_i+q_i=1$.

该链可以用如下图形表示:



由此可见, 状态 0 和 B 是 "吸收状态"; 对于任何其他状态 i, 质点以概率 r_i 滞留, 以概率 p_i 向右移动一步, 以概率 q_i 向左移动一步.

现在求①

$$\alpha(x)=\lim_{k\to\infty}\alpha_k(x),$$

① $\alpha_k(x)$ 是轨道在时间段 $[0,k]$ 上经点 A 处越出区间 (A,B) 的概率 (见 §9). —— 译者

即质点自点 x 出发早于到达状态 B 而到达状态 0 的极限概率. 当 $k\to\infty$ 时, 对于 $0<j<B$, 在 $\alpha_k(x)$ 的方程中求极限, 得

$$\alpha(j)=q_j\alpha(j-1)+r_j\alpha(j)+p_j\alpha(j+1),$$

且边界条件为

$$\alpha(0)=1,\quad \alpha(B)=0.$$

由于 $r_j=1-q_j-p_j$, 则

$$p_j[\alpha(j+1)-\alpha(j)]=q_j[\alpha(j)-\alpha(j-1)].$$

从而

$$\alpha(j+1)-\alpha(j)=\rho_j[\alpha(1)-1],$$

其中

$$\rho_j=\frac{q_1\cdots q_j}{p_1\cdots p_j},\quad \rho_0=1.$$

因为

$$\alpha(j+1)-1=\sum_{i=0}^{j}[\alpha(i+1)-\alpha(i)],$$

所以

$$\alpha(j+1)-1=[\alpha(1)-1]\sum_{i=0}^{j}\rho_i.$$

如果 $j=B-1$, 则 $\alpha(j+1)=\alpha(B)=0$, 即

$$\alpha(1)-1=-\frac{1}{\sum\limits_{i=0}^{B-1}\rho_i}.$$

于是,

$$\alpha(1)=\frac{\sum\limits_{i=1}^{B-1}\rho_i}{\sum\limits_{i=0}^{B-1}\rho_i},\quad \alpha(j)=\frac{\sum\limits_{i=j}^{B-1}\rho_i}{\sum\limits_{i=0}^{B-1}\rho_i},\quad j=1,\cdots,B.$$

(与 §9 中相应的结果比较.)

现在设

$$m(x)=\lim_{x\to\infty}m_k(x)$$

是游动的质点在到达状态 0 或 B 之一前, 所度过时间的极限值. 那么, $m(0)=m(B)=0$,

$$m(x)=1+\sum_y m(y)p_{xy},$$

从而, 对于所研究的例子, 对于一切 $j=1,\cdots,B-1$, 有

$$m(j)=1+q_j m(j-1)+r_j m(j)+p_j m(j+1).$$

为求 $m(j)$, 记

$$M(j)=m(j)-m(j-1),\quad j=1,\cdots,B.$$

那么,

$$p_j M(j+1)=q_j M(j)-1,\quad j=1,\cdots,B-1,$$

从而求得

$$M(j+1)=\rho_j M(1)-R_j,$$

其中

$$\rho_j=\frac{q_1\cdots q_j}{p_1\cdots p_j},\quad R_j=\frac{1}{p_j}\left[1+\frac{q_j}{p_{j-1}}+\cdots+\frac{q_j\cdots q_2}{p_{j-1}\cdots p_1}\right].$$

因此

$$\begin{aligned}m(j)=m(j)-m(0)&=\sum_{i=0}^{j-1}M(i+1)\\&=\sum_{i=0}^{j-1}[\rho_i m(1)-R_i]=m(1)\sum_{i=0}^{j-1}\rho_i-\sum_{i=0}^{j-1}R_i.\end{aligned}$$

最后, 只剩下求 $m(1)$. 由于 $m(B)=0$, 可见

$$m(1)=\frac{\sum\limits_{i=0}^{B-1}R_i}{\sum\limits_{i=0}^{B-1}\rho_i},$$

而对于 $1<j\leqslant B$,

$$m(j)=\sum_{i=0}^{j-1}\rho_i\times\frac{\sum\limits_{i=0}^{B-1}R_i}{\sum\limits_{i=0}^{B-1}\rho_i}-\sum_{i=0}^{j-1}R_i.$$

(与 §9 相应的结果比较, 其中 $r_i=0, p_i=p, q_i=q$).

6. 强马尔可夫性 这一小节研究马尔可夫性 (8) 的一种强化: 将普通时间 k 换成随机时间后马尔可夫性仍然成立 (见下面定理 2). 这一称为**强马尔可夫性**的概念的重要性, [例如在推导递推公式 (38) 的过程中] 将得到证实, 而公式 (38) 的递推公式对于马尔可夫链状态的分类至关重要 (第八章).

设 $\xi = (\xi_0, \cdots, \xi_n)$ 是齐次马尔可夫链, (p_{ij}) 是转移概率矩阵, $\mathscr{D}^\xi = (\mathscr{D}_k^\xi)_{0\leqslant k\leqslant n}$ 是分割系, 其中 $\mathscr{D}_k^\xi = \mathscr{D}_{\xi_0,\cdots,\xi_k}$. 以 $\mathscr{B}_k^\xi$ 表示由分割 $\mathscr{D}_k^\xi$ 生成的代数 $\alpha(\mathscr{D}_k^\xi)$.

我们首先给予马尔可夫性 (8) 另一种表达形式. 设 $B \in \mathscr{B}_k^\xi$, 并证明

$$\begin{aligned}&\mathbf{P}(\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1} | B \cap \{\xi_k = a_k\})\\ &= \mathbf{P}(\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1} | \xi_k = a_k)\end{aligned} \tag{29}$$

(假设 $\mathbf{P}(B \cap \{\xi_k = a_k\}) > 0$). 实际上, 集合 B 可以表示为

$$B = \sum\nolimits^* \{\xi_0 = a_0^*, \cdots, \xi_k = a_k^*\},$$

其中 $\sum^*$ 表示对某个数组 $(a_0^*, \cdots, a_k^*)$ 求和. 因此

$$\begin{aligned}&\mathbf{P}(\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1} | B \cap \{\xi_k = a_k\})\\ &= \frac{\mathbf{P}(\{\xi_n = a_n, \cdots, \xi_k = a_k\} \cap B)}{\mathbf{P}(\{\xi_k = a_k\} \cap B)}\\ &= \frac{\sum^* \mathbf{P}(\{\xi_n = a_n, \cdots, \xi_k = a_k\} \cap \{\xi_0 = a_0^*, \cdots, \xi_k = a_k^*\})}{\mathbf{P}(\{\xi_k = a_k\} \cap B)},\end{aligned} \tag{30}$$

而由于马尔可夫性, 可见

$$\begin{aligned}&\mathbf{P}(\{\xi_n = a_n, \cdots, \xi_k = a_k\} \cap \{\xi_0 = a_0^*, \cdots, \xi_k = a_k^*\})\\ &= \begin{cases}\mathbf{P}(\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1} | \xi_0 = a_0^*, \cdots, \xi_k = a_k^*) & \\ \qquad \times \mathbf{P}(\xi_0 = a_0^*, \cdots, \xi_k = a_k^*), & \text{若 } a_k = a_k^*,\\ 0, & \text{若 } a_k \neq a_k^*,\end{cases}\\ &= \begin{cases}\mathbf{P}(\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1} | \xi_k = a_k) & \\ \qquad \times \mathbf{P}(\xi_0 = a_0^*, \cdots, \xi_k = a_k^*), & \text{若 } a_k = a_k^*,\\ 0, & \text{若 } a_k \neq a_k^*,\end{cases}\\ &= \begin{cases}\mathbf{P}(\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1} | \xi_k = a_k) & \\ \qquad \times \mathbf{P}(\{\xi_k = a_k\} \cap B), & \text{若 } a_k = a_k^*,\\ 0, & \text{若 } a_k \neq a_k^*.\end{cases}\end{aligned}$$

从而 (30) 式中的和 $\sum^*$ 等于

$$\mathbf{P}(\xi_n = a_n, \cdots, \xi_{k+1} = a_{k+1} | \xi_k = a_k)\mathbf{P}(\{\xi_k = a_k\} \cap B),$$

于是,(29) 得证.

设 τ (关于分割系 $\mathscr{D}^\xi = (\mathscr{D}_k^\xi)_{0\leqslant k\leqslant n}$) 是停止时间 (见 §11 定义 2)

定义 称代数 $\mathscr{B}_n^\xi$ 中的集合 B 属于集合系 $\mathscr{B}_\tau^\xi$, 如果对于每一个 $1 \leqslant k \leqslant n$,

$$B \cap \{\tau = k\} \in \mathscr{B}_k^\xi \tag{31}$$

不难验证, 这样集合 B 的全体 $\mathscr{B}_\tau^\xi$ 构成代数 (称之为到时刻 τ 之前观测到的事件的代数).

定理 2 *设 $\xi = (\xi_0, \cdots, \xi_n)$ 是齐次马尔可夫链, (p_{ij}) 是转移概率矩阵, τ (关于分割系 $\mathscr{D}^\xi$) 是停止时间, $B \in \mathscr{B}_\tau^\xi$, 而 $A = \{\omega : \tau + l \leqslant n\}$. 那么, 如果*

$$\mathbf{P}(A \cap B \cap \{\xi_\tau = a_0\}) > 0,$$

则下面的强马尔可夫性成立:

$$\begin{aligned}
&\mathbf{P}(\xi_{\tau+l} = a_l, \cdots, \xi_{\tau+1} = a_1 | A \cap B \cap \{\xi_\tau = a_0\}) \\
&= \mathbf{P}(\xi_{\tau+l} = a_l, \cdots, \xi_{\tau+1} = a_1 | A \cap \{\xi_\tau = a_0\}),
\end{aligned} \tag{32}$$

且 (当 $\mathbf{P}(A \cap \{\xi_\tau = a_0\}) > 0$ 时)

$$\mathbf{P}(\xi_{\tau+l} = a_l, \cdots, \xi_{\tau+1} = a_1 | A \cap \{\xi_\tau = a_0\}) = p_{a_0 a_1} \cdots p_{a_{l-1} a_l}. \tag{33}$$

证明 为简便计, 我们只证明 $l = 1$ 的情形. 由于 $B \cap \{\tau = k\} \in \mathscr{B}_k^\xi$, 可见根据 (29) 式, 有

$$\begin{aligned}
&\mathbf{P}(\xi_{\tau+1} = a_1, A \cap B \cap \{\xi_\tau = a_0\}) \\
&= \sum_{k \leqslant n-1} \mathbf{P}\{\xi_{k+1} = a_1, \xi_k = a_0, \tau = k, B\} \\
&= \sum_{k \leqslant n-1} \mathbf{P}\{\xi_{k+1} = a_1 | \xi_k = a_0, \tau = k, B\} \mathbf{P}\{\xi_k = a_0, \tau = k, B\} \\
&= \sum_{k \leqslant n-1} \mathbf{P}\{\xi_{k+1} = a_1 | \xi_k = a_0\} \mathbf{P}\{\xi_k = a_0, \tau = k, B\} \\
&= p_{a_0 a_1} \sum_{k \leqslant n-1} \mathbf{P}\{\xi_k = a_0, \tau = k, B\} = p_{a_0 a_1} \mathbf{P}(A \cap B \cap \{\xi_\tau = a_0\}),
\end{aligned}$$

于是, (32) 和 (33) 式得证 (对于 (33) 式的情形, 应设 $B = \Omega$) □

注 1 显然, 对于 $l = 1$ 的情形, 强马尔可夫性 (32) 和 (33) 式等价于: 对于任意 $C \subseteq X$,

$$\mathbf{P}(\xi_{\tau+1} \in C | A \cap B \cap \{\xi_\tau = a_0\}) = \mathbf{P}_{a_0}(C), \tag{34}$$

其中

$$\mathbf{P}_{a_0}(C) = \sum_{a_1 \in C} p_{a_0 a_1}.$$

而 (34) 式可以表述为: 在集合 $A=\{\tau\leqslant n-1\}$ 上,

$$\mathbf{P}\{\xi_{\tau+1}\in C|\mathscr{B}_\tau^\xi\}=\mathbf{P}_{\xi_\tau}(C), \tag{35}$$

这在一般齐次马尔可夫过程论中, 是强马尔可夫性常用的形式之一

注 2 即便没有事件 $A\cap B\cap\{\xi_\tau=a_0\}$ 和 $A\cap\{\xi_\tau=a_0\}$ 的概率大于 0 的要求, 性质 (32) 和 (33) (只要利用注 1 所描绘的约定) 仍然成立.

7. n 步的转移概率 $p_{ij}^{(n)}$ 设 $\xi=(\xi_0,\cdots,\xi_n)$ 是齐次马尔可夫链, (p_{ij}) 是转移概率矩阵,

$$f_{ii}^{(k)}=\mathbf{P}(\xi_k=i;\xi_l\neq i,1\leqslant l\leqslant k-1|\xi_0=i), \tag{36}$$

而对于 $i\neq j$,

$$f_{ij}^{(k)}=\mathbf{P}(\xi_k=j,\xi_l\neq j,1\leqslant l\leqslant k-1|\xi_0=i), \tag{37}$$

分别是在时刻 k 首返状态 i 的概率和在时刻 k 首返状态 j 的概率.

现在证明

$$p_{ij}^{(n)}=\sum_{k=1}^n f_{ij}^{(k)}p_{jj}^{(n-k)},\ \text{其中}\ p_{jj}^{(0)}=1 \tag{38}$$

此式的直观意义很明显: 为经 n 步由状态 i 到达状态 j, 需要经 $k(1\leqslant k\leqslant n)$ 步首达状态 j, 然后经 $n-k$ 步由 j 返回 j. 现在给出 (38) 式的严格证明.

设 j 固定且

$$\tau_j=\min\{1\leqslant k\leqslant n:\xi_k=j\},$$

若 $\{\cdot\}=\varnothing$, 则设 $\tau_j=n+1$ 那么 $f_{ij}^{(k)}=\mathbf{P}(\tau_j=k|\xi_0=i)$, 而

$$\begin{aligned}p_{ij}^{(n)}&=\mathbf{P}(\xi_n=j|\xi_0=i)=\sum_{k=1}^n\mathbf{P}(\xi_n=j,\tau_j=k|\xi_0=i)\\&=\sum_{k=1}^n\mathbf{P}(\xi_{\tau_j+n-k}=j,\tau_j=k|\xi_0=i),\end{aligned} \tag{39}$$

其中最后一个等式成立, 因为在集合 $\{\tau_j=k\}$ 上 $\xi_{\tau_j+n-k}=\xi_n$. 而且, 对于任意 $1\leqslant k\leqslant n$, 集合 $\{\tau_j=k\}=\{\tau_j=k,\xi_{\tau_j}=j\}$ 因此, 如果 $\mathbf{P}\{\xi_0=i,\tau_j=k\}>0$, 则由定理 2, 有

$$\begin{aligned}\mathbf{P}(\xi_{\tau_j+n-k}=j|\xi_0=i,\tau_j=k)&=\mathbf{P}(\xi_{\tau_j+n-k}=j|\xi_0=i,\tau_j=k,\xi_{\tau_j}=j)\\&=\mathbf{P}(\xi_{\tau_j+n-k}=j|\xi_{\tau_j}=j)=p_{jj}^{(n-k)},\end{aligned}$$

而根据 (37) 式, 有

$$p_{ij}^{(n)}=\sum_{k=1}^n\mathbf{P}(\xi_{\tau_j+n-k}=j|\xi_0=i,\tau_j=k)\mathbf{P}(\tau_j=k|\xi_0=i)=\sum_{k=1}^n p_{jj}^{(n-k)}f_{ij}^{(k)},$$

于是, 关系 (38) 式得证

8. 练习题

1. 设 $\xi=(\xi_0,\cdots,\xi_n)$ 是马尔可夫链, X 是其值域, $f=f(x), x\in X$ 是某一函数. 问序列 $(f(\xi_0),\cdots,f(\xi_n))$ 是不是马尔可夫链? 序列 $(\xi_n,\xi_{n-1},\cdots,\xi_0)$ 是不是 "后向" 马尔可夫链?

2. 设 $\mathbb{P}=(p_{ij}), 1\leqslant i,j\leqslant r$ 是随机矩阵, 而 λ 是其特征值, 即特征方程 $\det(\mathbb{P}-\lambda E)=0$ 的根. 证明 $\lambda_1=1$ 是特征值, 而一切其他特征值 $\lambda_2,\cdots,\lambda_r$ 的绝对值都不大于 1. 如果所有特征值 $\lambda_1,\cdots,\lambda_r$ 两两不同, 则 $p_{ij}^{(k)}$ 有如下表示式:

$$p_{ij}^{(k)}=\pi_j+\alpha_{ij}(2)\lambda_2^k+\cdots+\alpha_{ij}(r)\lambda_r^k,$$

其中 $\pi_j,\alpha_{ij}(2),\cdots,\alpha_{ij}(r)$ 可以通过矩阵 $\mathbb{P}$ 的元素表示. (特别, 用这一代数方法分析马尔可夫链的性质, 可见当 $|\lambda_2|<1,\cdots,|\lambda_r|<1$ 时, 对于每一个 j, 存在不依赖于 i 的极限 $\lim_k p_{ij}^{(k)}$.)

3. 设 $\xi=(\xi_0,\cdots,\xi_n)$ 是齐次马尔可夫链, X 是其状态集, (p_{ij}) 是转移概率矩阵, 记

$$T\varphi(x)=\mathbf{E}[\varphi(\xi_1)|\xi_0=x]\left(=\sum_y\varphi(y)p_{xy}\right),$$

假设非负函数 $\varphi=\varphi(x)$ 满足方程

$$T\varphi(x)=\varphi(x),\quad x\in X.$$

证明随机变量序列

$$\zeta=(\zeta_k,\mathscr{D}_k^{\xi}),\quad 其中\ \zeta_k=\varphi(\xi_k)$$

是鞅

4. 设 $\xi=(\xi_n,\Pi,\mathbb{P})$ 和 $\widetilde{\xi}=(\widetilde{\xi}_n,\widetilde{\Pi},\mathbb{P})$ 是两个马尔可夫链, 具有不同的初始分布, 相应为 $\Pi=(p_1,\cdots,p_r)$ 和 $\widetilde{\Pi}=(\widetilde{p}_1,\cdots,\widetilde{p}_r)$. 设 $\Pi^{(n)}=(p_1^{(n)},\cdots,p_r^{(n)})$ 和 $\widetilde{\Pi}^{(n)}=(\widetilde{p}_1^{(n)},\cdots,\widetilde{p}_r^{(n)})$, 证明

$$\sum_{i=1}^r|\widetilde{p}_i^{(n)}-p_i^{(n)}|\leqslant 2(1-r\varepsilon)^n.$$

5. 设 P 和 Q 为随机矩阵. 证明 PQ 和 $\alpha P+(1-\alpha)Q, 0\leqslant\alpha\leqslant 1$ 也是随机矩阵.

6. 考虑齐次马尔可夫链 $(\xi_0,\cdots,\xi_n)$, $X=\{0,1\}$ 是其状态集, 其转移概率矩阵为

$$\begin{pmatrix}1-p & p\\ q & 1-q\end{pmatrix},$$

其中 $0<p<1, 0<q<1$. 设 $S_n=\xi_0+\cdots+\xi_n$, 作为 (§6) 棣莫弗–拉普拉斯定理

的推广, 证明

$$\mathbf{P}\left\{\frac{S_n-\dfrac{p}{p+q}n}{\sqrt{\dfrac{npq(2-p-q)}{(p+q)^3}}}\leqslant x\right\}\to\Phi(x),\qquad n\to\infty.$$

说明, 当 $p+q=1$ 时, 随机变量 $\xi_0,\cdots,\xi_n$ 独立, 而上述命题归结为:

$$\mathbf{P}\left\{\frac{S_n-pn}{\sqrt{npq}}\leqslant x\right\}\to\Phi(x),\qquad n\to\infty.$$

第二章　概率论的数学基础

像几何学和代数学一样, 概率论作为数学学科, 可以并且应该完全公理化. 这意味着, 在给出研究对象及其关系之后, 还应给出这些关系应服从的公理, 在此之后全部的叙述应仅仅以这些公理为基础, 而不依赖于这些对象及其关系的一般的具体值.

A. H. 柯尔莫戈洛夫,《概率论的基本概念》[32]

---

§1. 有无限种结局试验的概率模型. 柯尔莫戈洛夫公理化体系

1. 代数和有限 – 可加测度 上一章介绍的模型, 提供了具有*有限*种结局的试验的概率 – 统计描述. 例如, "三对象"

$$(\Omega, \mathscr{A}, \mathbf{P}), \text{ 其中 } \Omega = \{\omega : \omega = (a_1, \cdots, a_n), a_i = 0, 1\},$$
$$\mathscr{A} = \{A : A \subseteq \Omega\}, \mathbf{P}(\{\omega\}) = p^{\Sigma a_i}(1-p)^{n-\Sigma a_i}(= p(\omega)),$$

是 n 次 "独立" 地掷硬币模型, 其中 "正面" 出现的概率为 p. 在此模型中, 一切结局的个数 $N(\Omega)$, 即集合中点的个数是有限的, 并且等于 2^n.

现在的问题是, 建立*无限次* "独立地" 掷硬币模型, 其中每次掷 "正面" 出现的概率为 p.

作为结局的全体, 自然取集合

$$\Omega = \{\omega : \omega = (a_1, a_2, \cdots), a_i = 0, 1\},$$

即一切序列 $\omega = (a_1, a_2, \cdots)$ 的空间, 集合 ω 的元素有 0 和 1 两个可能值.

问集合 Ω 的势 (基数) $N(\Omega)$ 如何? 熟知, 任意实数 $a \in [0, 1)$ 可以唯一地分解为 (可能含无限个 0 的) 二进制小数:

$$a = \frac{a_1}{2} + \frac{a_2}{2^2} + \cdots (a_i = 0, 1).$$

由此可见, 在 ω 的集合 Ω 与 a 的集合 [0,1) 之间存在一一对应关系, 说明集合 Ω 的势等于*连续统*的势.

于是, 如果希望建立描绘 "*无限次* '独立地' 掷硬币类型" 的试验的概率模型, 则不得不考虑相当复杂来源的空间 Ω.

我们现在尝试弄清楚, 如何在 "无限次 '独立地' 掷 (对称 [$p = q = 1/2$] 的) 硬币" 的模型中, 合理地确定 (赋予) 概率.

由于作为 Ω 可以取集合 [0,1), 所以把上面提出的问题, 可以视为 "在集合 [0, 1) 上随机选点" 的模型中求概率的问题. 由于直观的对称性, 可以认为一切结局都是

"等可能的". 由于集合 $[0,1)$ 是不可数的, 假如考虑到点属于集合 $[0,1)$ 的概率应该等于 1, 则每个 ω 的概率必为 0. 但是如此定义概率 $(p(\omega)=0, \omega \in [0,1))$ 没有什么意义. 然而, 我们通常关心的并不是个别结局出现的概率, 而是试验结局属于某个给定集合 (事件) A 的概率. 在初等概率论中, 根据 "权重" $p(\omega)$ 即可求事件 A 的概率:

$$\mathbf{P}(A)=\sum_{\omega \in A} p(\omega).$$

对于现在考虑的情况, 例如, 当 $p(\omega)=0, \omega \in [0,1)$ 时, 我们无法求 "自 $[0,1)$ 上随机选的点" 属于集合 $[0,1/2)$ 的概率. 然而, 直观上很明显这一概率应该等于 1/2.

这种情况提示我们, 当空间 Ω 不可数时, 不应求个别结局的概率, 而应求 Ω 中某些集合的概率. 第一章的论证说明, 在其上确定概率的集合的全体, 关于并、交与补应该是封闭的. 鉴于这种情况, 我们引进如下定义.

定义 1 设 Ω 是点 ω 的某一集合. Ω 的子集系 $\mathscr{A}$ 称做代数, 如果

a) $\Omega \in \mathscr{A}$,

b) $A, B \in \mathscr{A} \Rightarrow A \bigcup B \in \mathscr{A}, A \bigcap B \in \mathscr{A}$,

c) $A \in \mathscr{A} \Rightarrow \overline{A} \in \mathscr{A}$.

注意, 在条件 b) 中, 只需要求 $A \bigcup B \in \mathscr{A}$ 或 $A \bigcap B \in \mathscr{A}$ 之一成立, 因为

$$A \bigcup B=\overline{\overline{A} \bigcap \overline{B}} \quad A \bigcap B=\overline{\overline{A} \bigcup \overline{B}}$$

为表述概率模型的概念, 需要下面的概念.

定义 2 设 $\mathscr{A}$ 是 Ω 的子集的代数. 在 $[0,\infty]$ 上取值的集函数 $\mu=\mu(A), A \in \mathscr{A}$, 称做定义在 $\mathscr{A}$ 上的有限 - 可加测度, 如果 $\mathscr{A}$ 中的任意两个不相交集合 A 和 B, 有

$$\mu(A+B)=\mu(A)+\mu(B). \tag{1}$$

具有 $\mu(\Omega)<\infty$ 的有限 - 可加测度 $\mu=\mu(A)$ 称做有限的, 而当 $\mu(\Omega)=1$ 时, 称做有限 - 可加概率测度或有限 - 可加概率.

2. 概率模型 现在给出试验的概率模型, 假设其 (广义) 结局 (现象) 属于集合 Ω.

定义 3 三对象

$$(\Omega, \mathscr{A}, \mathbf{P}),$$

的总体, 其中

a) Ω 是点 ω 的集合,

b) $\mathscr{A}$ 是 Ω 的子集的代数,

c) $\mathbf{P}$ 是 $\mathscr{A}$ 上的有限 - 可加概率,

称做概率模型、广义 (试验的) 概率 "理论".

不过, 事实表明, 为建立有成效的数学理论, 这一概率模型显得过于广泛. 因此, 不但对所考虑的 Ω 的子集类, 而且对容许的概率测度类, 不得不加以限制.

定义 4 集合 Ω 的子集系 $\mathscr{F}$ 称做 $\sigma-$ 代数, 如果 $\mathscr{F}$ 是代数, 而且满足下列性质 (定义 1 之性质 b) 的加强):

b*) 如果 $A_n \in \mathscr{F}, n = 1, 2, \cdots$, 则

$$\bigcup A_n \in \mathscr{F}, \bigcap A_n \in \mathscr{F}$$

(这时, 只需要求二式之一成立).

定义 5 空间 Ω 连同其子集的 $\sigma-$ 代数 $\mathscr{F}$, 称做可测空间, 记作 $(\Omega, \mathscr{F})$.

定义 6 定义在集合 Ω 的子集的代数 $\mathscr{A}$ 上的有限 – 可加测度 μ, 称做完全可加或 $\sigma-$ 可加测度, 亦简称测度, 如果对于两两不相交的集合 $A_1, A_2, \cdots \in \mathscr{A}$, 且

$$\sum_{n=1}^{\infty} A_n \in \mathscr{A},$$

有

$$\mu\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

测度 μ 称做 $\sigma-$ 有限的, 如果空间 Ω 可以表示为

$$\Omega = \sum_{n=1}^{\infty} \Omega_n, \Omega_n \in \mathscr{A},$$

且 $\mu(\Omega_n) < \infty, n = 1, 2, \cdots$.

代数 $\mathscr{A}$ 上的测度 (注意, 指 $\sigma-$ 可加测度) $\mathbf{P}$, 如果满足条件 $\mathbf{P}(\Omega) = 1$, 则 $\mathbf{P}$ 称做 (定义在代数 $\mathscr{A}$ 的集合上的) 概率测度或简称概率.

现在指出概率测度的某些性质.

1° 若 $\varnothing$ 是空集, 则

$$\mathbf{P}(\varnothing) = 0.$$

2° 若 $A, B \in \mathscr{A}$, 则

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

3° 若 $A, B \in \mathscr{A}$ 且 $B \subseteq A$, 则

$$\mathbf{P}(B) \leqslant \mathbf{P}(A).$$

4° 若 $A_n \in \mathscr{A}, n = 1, 2, \cdots$, 且 $\bigcup A_n \in \mathscr{A}$, 则

$$\mathbf{P}(A_1 \cup A_2 \cup \cdots) \leqslant \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots.$$

前三条性质显然. 为证明性质 4, 只需注意到, 对于 $n \geqslant 2$, 有

$$\bigcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} B_n, B_1 = A_1, B_n = \left(\bigcap_{k=1}^{n-1} A_k\right) \cap A_n, B_i \cap B_j = \varnothing,$$

因而

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbf{P}\left(\sum_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbf{P}(B_n) \leqslant \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

下面的定理证明, 集合的有限 – 可加函数, 同时也是 $\sigma-$ 可加的. 该定理有许多应用.

定理 设 $\mathbf{P}$ 是定义在代数 $\mathscr{A}$ 上的, $\sigma-$ 可加集合函数, 且 $\mathbf{P}(\Omega) = 1$. 那么, 如下 4 个条件等价:

1) $\mathbf{P}$ 为 $\sigma-$ 可加 ($\mathbf{P}$ 是概率);

2) $\mathbf{P}$ 上连续, 即对于任意集合 $A_1, A_2, \cdots \in \mathscr{A}$, 若

$$A_n \subseteq A_{n+1}, \bigcup_{n=1}^{\infty} A_n \in \mathscr{A},$$

则

$$\lim_n \mathbf{P}(A_n) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right);$$

3) $\mathbf{P}$ 下连续, 即对于任意集合 $A_1, A_2, \cdots \in \mathscr{A}$, 若

$$A_n \supseteq A_{n+1}, \bigcap_{n=1}^{\infty} A_n \in \mathscr{A},$$

则

$$\lim_n \mathbf{P}(A_n) = \mathbf{P}\left(\bigcap_{n=1}^{\infty} A_n\right);$$

4) $\mathbf{P}$ 在 "0" 连续, 即对于任意集合 $A_1, A_2, \cdots \in \mathscr{A}$, 若

$$A_{n+1} \subseteq A_n, \bigcap_{n=1}^{\infty} A_n = \varnothing,$$

则

$$\lim_n \mathbf{P}(A_n) = 0.$$

证明 1)⇒2). 由于

$$\bigcup_{n=1}^{\infty} A_n = A_1 + (A_2 \backslash A_1) + (A_3 \backslash A_2) + \cdots,$$

则

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbf{P}(A_1) + \mathbf{P}(A_2 \backslash A_1) + \mathbf{P}(A_3 \backslash A_2) + \cdots$$
$$= \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1) + \mathbf{P}(A_3) - \mathbf{P}(A_2) + \cdots$$
$$\fallingdotseq \lim_n \mathbf{P}(A_n).$$

2)⇒3). 设 $n \geqslant 1$, 则

$$\mathbf{P}(A_n) = \mathbf{P}(A_1 \backslash (A_1 \backslash A_n)) = \mathbf{P}(A_1) - \mathbf{P}(A_1 \backslash A_n).$$

集合序列 $(A_1 \backslash A_n)_{n\geqslant 1}$ 是非降的 (见 §1 表 1), 而

$$\bigcup_{n=1}^{\infty} (A_1 \backslash A_n) = A_1 \backslash \bigcap_{n=1}^{\infty} A_n.$$

那么, 由 2) 可见

$$\lim_n \mathbf{P}(A_1 \backslash A_n) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} (A_1 \backslash A_n)\right),$$

因此

$$\lim_n \mathbf{P}(A_n) = \mathbf{P}(A_1) - \lim_n \mathbf{P}(A_1 \backslash A_n)$$
$$= \mathbf{P}(A_1) - \mathbf{P}\left(\bigcup_{n=1}^{\infty} (A_1 \backslash A_n)\right) = \mathbf{P}(A_1) - \mathbf{P}\left(A_1 \backslash \bigcap_{n=1}^{n} A_n\right)$$
$$= \mathbf{P}(A_1) - \mathbf{P}(A_1) + \mathbf{P}\left(\bigcap_{n=1}^{n} A_n\right) = \mathbf{P}\left(\bigcap_{n=1}^{n} A_n\right).$$

3)⇒4). 显然.

4)⇒1). 设两两不相交的集合 $A_1, A_2, \cdots \in \mathscr{A}$, 且 $\sum_{n=1}^{\infty} A_n \in \mathscr{A}$, 则

$$\mathbf{P}\left(\sum_{i=1}^{\infty} A_i\right) = \mathbf{P}\left(\sum_{i=1}^{n} A_i\right) + \mathbf{P}\left(\sum_{i=n+1}^{\infty} A_i\right),$$

而由于

$$\sum_{i=n+1}^{\infty} A_i \downarrow \varnothing, n \to \infty,$$

因此,

$$\sum_{i=1}^{\infty} \mathbf{P}(A_i) = \lim_n \sum_{i=1}^{n} \mathbf{P}(A_i) = \lim_n \mathbf{P}\left(\sum_{i=1}^{n} A_i\right)$$
$$= \lim_n \left[\mathbf{P}\left(\sum_{i=1}^{\infty} A_i\right) - \mathbf{P}\left(\sum_{i=n+1}^{\infty} A_i\right)\right]$$
$$= \mathbf{P}\left(\sum_{i=1}^{\infty} A_i\right) - \lim_n \mathbf{P}\left(\sum_{i=n+1}^{\infty} A_i\right) = \mathbf{P}\left(\sum_{i=1}^{\infty} A_i\right). \quad \square$$

3. 柯尔莫戈洛夫公理化体系 设 Ω 是试验 E 的结局 (现象) 的集合. 现在可以表述柯尔莫戈洛夫公理化体系, 该公理化体系已经被普遍地接受, 并且是建立试验 E 的概率模型的基础.

基本定义 考虑三个对象的全体

$$(\Omega, \mathscr{F}, \mathbf{P}),$$

其中

a) Ω 是点 ω 的集合,

b) $\mathscr{F}$ 是 Ω 子集的 $\sigma-$代数,

c) $\mathbf{P}$ 是 $\mathscr{F}$ 上的概率,

称 $(\Omega, \mathscr{F}, \mathbf{P})$ 为 (试验的) 概率模型或概率空间. 这时, 结局的空间 Ω 称为基本事件空间, $\mathscr{F}$ 中的集合 A 称做事件, 而 $\mathbf{P}(A)$ 称做事件 A 的概率.

由这一定义可见, 概率的公理本质上依赖于集合论与测度论的工具. 因此, 下面的表 2–1 很重要, 该表将集合论与概率论相应的概念进行对照和比较. 在下面的两节将陆续给出概率论中最重要的可测空间的例子, 以及其中确定概率方法的例子.

表 2 – 1

记 号	集 合 论	概 率 论
ω	元素, 点	结局, 基本事件
Ω	点集	结局的空间, 基本事件空间; 必然事件
$\mathscr{F}$	子集的 $\sigma-$ 代数	事件的 $\sigma-$ 代数
$A \in \mathscr{F}$	点集	事件 (若 $\omega \in A$, 则称事件 A 出现了)
$\overline{A} = \Omega \backslash A$	集合 A 的补集: 不属于 A 的点 ω 的集合	事件 "事件 A 不出现"
$A \cup B$	集合 A 与 B 的并; 属于 A 或 B 的点 ω 的集合	事件 "事件 A 或 B 至少一个出现"
$A \cap B$ 或 AB	集合 A 与 B 的交: 同属于 A 和 B 的点 ω 的集合	事件 "事件 A 或 B 同时出现"
$\varnothing$	空集	不可能事件
$A \cap B = \varnothing$	集合 A 与 B 不相交	事件 A 与 B 不相容 (不能同时出现)
$A + B$	集合 A 与 B 之和: 即不相交集合 A 与 B 的并	事件 "出现了二不相容事件之一"
$A \backslash B$	集合 A 与 B 之差: 属于 A 不属于 B 的点ω的集合	事件 "A 出现但 B 不出现"
$A \triangle B$	集合的对称差: 集合 $(A \backslash B) \cup (B \backslash A)$	事件 " A 或 B 仅出现一个"
$\bigcup\limits_{n=1}^{\infty} A_n$	集合 $A_1, A_2, \cdots$ 的并: 至少属于 $A_1, A_2, \cdots$ 之一的点 ω 的集合	事件 "$A_1, A_2, \cdots$ 的并,": 事件 "$A_1, A_2, \cdots$ 至少出现一个"
$\sum\limits_{n=1}^{\infty} A_n$	集合 $A_1, A_2, \cdots$ 的和: 两两不交集合 $A_1, A_2, \cdots$ 的并	事件"两两不相容事件 $A_1, A_2, \cdots$ 中至少出现一个"
$\bigcap\limits_{n=1}^{\infty} A_n$	集合 $A_1, A_2, \cdots$ 的交: 同属于 $A_1, A_2, \cdots$ 的点 ω 的集合	事件 "$A_1, A_2, \cdots$ 的交" : 事件 "$A_1, A_2, \cdots$ 同时出现"
$A_n \uparrow A$ 或 $\mathrm{A} = \lim\limits_{n} \uparrow A_n$	递增集合序列 $\{A_n\}$ 收敛于集合 A : $A_1 \subseteq A_2 \subseteq \cdots, A = \bigcup\limits_{n=1}^{\infty} A_n$	递增事件序列 $A_1, A_2, \cdots$ 收敛于事件 A

续表

记号	集合论	概率论
$A_n \downarrow A$ 或 $A = \lim\limits_n \downarrow A_n$	递减集合序列 $\{A_n\}$ 收敛于集合 A: $A_1 \supseteq A_2 \supseteq \cdots, A = \bigcap\limits_{n=1}^{\infty} A_n$	递减事件序列 $A_1, A_2, \cdots$ 收敛于事件 A
$\overline{\lim\limits_n} A_n$ 或 $\limsup A_n$ 或 {无限多个 A_n}	集合 $A = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$	在事件序列 $A_1, A_2, \cdots$ 中出现无限多次的 ω 的集合
$\underline{\lim\limits_n} A_n$ 或 $\liminf A_n$	集合 $A = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} A_k$	事件 "在 $A_1, A_2, \cdots$ 中出现无限多个事件, 其中仅有限个事件可能除外"

注 在建立用于描绘 "现象与条件" 的概率联系的试验模型时, 应 (根据第一章 §1 的注释) 事先说明, 是 "在何条件组合下" 考虑有关试验的. 通常我们并不作这种说明, 不过每次都应当清楚, 这些 "条件" 是什么.

4. 练习题

1. 设 $\Omega = \{r : r \in [0,1]\}$ 是区间 [0,1] 上有理点的集合, $\mathscr{A}$ 是集合代数: 其中每一个都是有限个形如 $\{r : a < r < b\}, \{r : a \leqslant r < b\}, \{r : a < r \leqslant b\}, \{r : a \leqslant r \leqslant b\}$ 的不相交集合 A 之并, 而 $\mathbf{P}(A) = b - a$. 证明, $\mathbf{P}(A), A \in \mathscr{A}$, 是有限 – 可加的集函数, 但不是 $\sigma-$ 可加的集函数.

2. 设 Ω 是可数集合, 而 $\mathscr{F}$ 是一切子集的全体. 设 $\mu(A) = 0$, 如果 A 有限, 而 $\mu(A) = \infty$, 如果 A 无限. 证明集函数 μ 有限 – 可加, 但非 $\sigma-$ 可加.

3. 设 μ 是 $\sigma-$ 代数 $\mathscr{F}$ 上的有限测度, $A_n \in F, n = 1, 2, \cdots$, 而

$$A = \lim_n A_n \left(\text{即 } A = \overline{\lim_n} A_n = \underline{\lim_n} A_n \right).$$

证明,$\mu(A) = \lim\limits_n \mu(A_n)$.

4. 证明, $\mathbf{P}(A \triangle B) = \mathbf{P}(A) + \mathbf{P}(B) - 2\mathbf{P}(A \bigcap B)$. (对照第一章 §1 练习题 4.)

5. 设

$$\rho_1(A, B) = \mathbf{P}(A \triangle B),$$
$$\rho_2(A, B) = \begin{cases} \dfrac{\mathbf{P}(A \triangle B)}{\mathbf{P}(A \cup B)}, & \text{若 } \mathbf{P}(A \cup B) \neq 0, \\ 0, & \text{若 } \mathbf{P}(A \cup B) = 0. \end{cases}$$

证明, $\rho_1(A, B)$ 和 $\rho_2(A, B)$ 可以作为 "距离", 即满足 "三角形不等式".

6. 设 μ 是代数 $\mathscr{A}$ 上的有限 – 可加测度, $A_1, A_2, \cdots \in \mathscr{A}$ 是两两不相交集合,

$$A = \sum_{i=1}^{n} A_n \in \mathscr{A}$$

证明

$$\mu(A) \geqslant \sum_{i=1}^{\infty} \mu(A_i).$$

7. 证明

$$\overline{\limsup A_n} = \liminf \overline{A}_n, \quad \overline{\liminf A_n} = \limsup \overline{A}_n,$$
$$\liminf A_n \subseteq \limsup A_n, \quad \limsup(A_n \cup B_n) = \limsup A_n \cup \limsup B_n,$$
$$\liminf(A_n \cap B_n) = \liminf A_n \cap \liminf B_n,$$
$$\limsup A_n \cap \liminf B_n \subseteq \limsup(A_n \cap B_n) \subseteq \limsup A_n \cap \limsup B_n.$$

如果 $A_n \uparrow A$ 或 $A_n \downarrow A$, 则

$$\lim\inf A_n = \limsup A_n.$$

8. 设 $\{x_n\}$ 是一数列, 而 $A_n = (-\infty, x_n)$. 证明, $x = \limsup x_n$ 与 $A = \limsup A_n$ 有如下联系: $(-\infty, x) \subseteq A \subseteq (-\infty, x]$. 换句话说, A 等于 $(-\infty, x)$ 或 $(-\infty, x]$.

9. 举例说明, 对于取 $+\infty$ 为值的测度, 由完全可加性, 一般推不出在 "零" $\varnothing$ 的连续性.

10. 证明布尔 (G. Boole) 不等式: $\mathbf{P}(A \bigcap B) \geqslant 1 - \mathbf{P}(\overline{A}) - \mathbf{P}(\overline{B})$.

11. 设 $A_1, \cdots, A_n$ 是 $\mathscr{F}$ 中的事件. 称这一事件组为**可重置的或可交换的** (exchangeable 或 interchangeable), 如果对于一切 $1 \leqslant k \leqslant n$ 和任意下标 $1 \leqslant i_1 < \cdots < i_k \leqslant n$, 概率 $\mathbf{P}(A_{i_1}, \cdots, A_{i_k})$ 等于同一个值 ($=p_k$). 对这样的事件, 证明下面的公式:

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = np_1 - \mathrm{C}_n^2 p_2 + \mathrm{C}_n^3 p_3 - \cdots + (-1)^{n-1} p_n.$$

12. 设 $(A_k)_{k\geqslant 1}$ 是无限可重置事件序列, 即对于一切 $n \geqslant 1$ 和任意下标 $1 \leqslant i_1 < \cdots < i_n$, 概率 $\mathbf{P}(A_{i_1}, \cdots, A_{i_n})$ 等于同一个值 ($= p_n$). 证明

$$\mathbf{P}\left(\underline{\lim_n} A_n\right) = \mathbf{P}\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{j\to\infty} p_j,$$
$$\mathbf{P}\left(\overline{\lim_n} A_n\right) = \mathbf{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = 1 - \lim_{j\to\infty} (-1)^j \Delta^j(p_0),$$

其中 $p_0 = 1, \Delta^1(p_n) = p_{n+1} - p_n, \Delta^j(p_n) = \Delta^1(\Delta^{j-1}(p_n)), j \geqslant 2$.

13. 设 $(A_n)_{n\geqslant 1}$ 是一集合序列, $I(A_n)$ 是集合 $A_n, n \geqslant 1$ 的示性函数. 证明

$$I\left(\underline{\lim_n} A_n\right) = \underline{\lim_n} I(A_n), I\left(\overline{\lim_n} A_n\right) = \overline{\lim_n} I(A_n),$$
$$I\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \sum_{n=1}^{\infty} I(A_n).$$

14. 证明

$$I\left(\bigcup_{n=1}^{\infty} A_n\right) = \max_{n\geqslant 1} I(A_n), I\left(\bigcap_{n=1}^{\infty} A_n\right) = \min_{n\geqslant 1} I(A_n).$$

15. 证明

$$\mathbf{P}\left(\overline{\lim_n} A_n\right) \geqslant \overline{\lim_n} \mathbf{P}(A_n), \mathbf{P}\left(\underline{\lim_n} A_n\right) \leqslant \underline{\lim_n} \mathbf{P}(A_n).$$

16. 设 $A^* = \overline{\lim}_n A_n$ 和 $A_* = \underline{\lim}_n A_n$, 证明$\mathbf{P}(A_n - A_*) \to 0, \mathbf{P}(A^* - A_n) \to 0$.

17. 在上一题的记号下, 记 $A = A^* = A_*$. 证明, 若 $A_n \to A$, 则 $\mathbf{P}(A \triangle A_n) \to 0$.

18. 设集合 A_n 在 $\mathbf{P}(A \triangle A^*) = \mathbf{P}(A \triangle A_*) = 0$ 意义下收敛于集合 A, 证明 $\mathbf{P}(A \triangle A_n) \to 0$.

19. 证明, 集合 A 与 B 的对称差 $A \triangle B$, 具有下面的性质:

$$I(A \triangle B) = I(A) + I(B)(\mathrm{mod} 2).$$

(由此可以导出 $\mathbf{P}(A \triangle B) = \mathbf{P}(A) + \mathbf{P}(B) - 2\mathbf{P}(A \bigcap B)$; 见练习题 4.) 证明, 对称差的下列性质:

$$(A \triangle B) \triangle C = A \triangle (B \triangle C), (A \triangle B) \triangle (B \triangle C) = A \triangle C,$$

$$A \triangle B = C \Leftrightarrow A = B \triangle C.$$

§2. 代数和 $\sigma-$ 代数. 可测空间.

1. 代数和 $\sigma-$ 代数 在建立 (试验的) 概率空间时, 代数和 $\sigma-$ 代数是组成元素. 关于代数和 $\sigma-$ 代数, 将举一些例子并介绍一系列性质.

设 Ω 是一基本事件空间. 集合系

$$\mathscr{F}_* = \{\varnothing, \Omega\}, \mathscr{F}^* = \{A : A \subseteq \Omega\}$$

显然是代数, 同时也是 $\sigma-$ 代数. 这时, $\mathscr{F}_*$ 是平凡的、最 "贫乏" 的 $\sigma-$ 代数, 而 $\mathscr{F}^*$ 是由 Ω 的一切子集构成的、最 "广泛" 的 $\sigma-$ 代数.

对于有限空间 Ω, $\sigma-$ 代数 $\mathscr{F}^*$ 非常直观, 一般这正是在初等概率论里所考虑的 "事件" 系. 对于不可数空间 Ω , 集系 $\mathscr{F}^*$ 显得过于广泛, 致使在这样的集系上不是总能 "合适地" 定义概率.

如果 $A \subseteq \Omega$, 则集系

$$\mathscr{F}_A = \{A, \overline{A}, \varnothing, \Omega\}$$

也是代数 ($\sigma-$ 代数) 的例, 并称做**由集合 A 诱导的**或**由集合 A 生成的代数** ($\sigma-$ 代数).

这一集系是由**分割**生成的集系的特殊情形. 确切地说, 设

$$\mathscr{D} = \{D_1, D_2, \cdots\}$$

是某一把 Ω 分为非空集合的有限**分割**:

$$\Omega = D_1 + D_2 + \cdots ; D_i \cap D_j = \varnothing, i \neq j.$$

那么, 由分割的有限或可数个元素之并形成的集系 $\mathscr{A}=\alpha(\mathscr{D})$, (连同空集) 是代数 ($\sigma-$ 代数).

下面的引理有重要意义, 因为它证明 "建立含给定集系的最小代数 ($\sigma-$ 代数)" 原则上是可能的.

引理 1 设 $\mathscr{E}$ 是 Ω 中的一个集系. 那么, 存在含 $\mathscr{E}$ 中所有集合的最小代数 $\alpha(\mathscr{E})$ 和最小 $\sigma-$ 代数 $\sigma(\mathscr{E})$.

证明 Ω 中一切子集的集系 $\mathscr{F}^*$ 是 $\sigma-$ 代数. 因此, 至少存在一个包含 $\mathscr{E}$ 的代数和一个包含 $\mathscr{E}$ 的 $\sigma-$ 代数. 现在, 由属于包含的 $\mathscr{E}$ 的任意代数 ($\sigma-$ 代数) 的一切集合, 组成集系 $\alpha(\mathscr{E})(\sigma(\mathscr{E}))$. 不难验证, 代数 $\alpha(\mathscr{E})$ ($\sigma-$ 代数 $\sigma(\mathscr{E})$) 就是引理所要求的最小代数 ($\sigma-$ 代数).

注 1 常把集系 $\alpha(\mathscr{E})$ (相应的 $\sigma(\mathscr{E})$) 称做由集系 $\mathscr{E}$ 诱导的 (生成的) 最小代数 (相应的最小 $\sigma-$ 代数).

前面已经指出 (§1 第 3 小节), $\sigma-$ 代数在概率空间的定义中占重要地位. 因此, 希望得到构造由某一代数 $\mathscr{A}$ 生成的 $\sigma-$ 代数的构造方法 (引理 1 仅给出了这样 $\sigma-$ 代数的存在性, 但没有提供其有效的构造方法).

下面是可以想到的且很自然的, 由 $\mathscr{A}$ 构造 $\sigma-$ 代数 $\sigma(\mathscr{A})$ 的方法之一.

设 $\mathscr{E}$ 是 Ω 的一个子集系. 设 $\widehat{\mathscr{E}}$ 是 Ω 的一个子集系, 包含 $\mathscr{E}$ 中的一切子集及其补集, 以及 $\mathscr{E}$ 中集合的有限或可数个集合的并. 令 $\mathscr{A}_0=\mathscr{A},\mathscr{A}_1=\widehat{\mathscr{A}_0},\mathscr{A}_2=\widehat{\mathscr{A}_1},\cdots$. 显然对于每个 n, 集合 $\mathscr{A}_n$ 属于 $\sigma(\mathscr{A})$, 而且似乎可以指望, 对于某个 n 有 $\mathscr{A}_n=\sigma(\mathscr{A})$, 或者至少 $\bigcup\mathscr{A}_n=\sigma(\mathscr{A})$.

然而, 事情一般并非如此. 事实上, 取 $\Omega=(0,1]$, 而作为代数 $\mathscr{A}$, 考虑由空集 $\varnothing$ 以及形如 $(a,b]$ 的区间之有限和诱导的 Ω 的子集系, 其中 a,b 是有理数. 不难证明, 这时集系 $\bigcup_{n=1}^{\infty}\mathscr{A}_n$ 严格小于 $\sigma-$ 代数 $\sigma(\mathscr{A})$.

我们以后主要关心的, 不是如何由代数 $\mathscr{A}$ 构造最小 $\sigma-$ 代数 $\sigma(\mathscr{A})$ 的问题, 而是确定给定的不同集系是否 $\sigma-$ 代数 $\sigma(\mathscr{A})$ 的问题.

为得到上述问题的答案, 要用到如下重要的概念 —— "单调类".

定义 1 Ω 的子集系 $\mathscr{M}$, 称做单调类, 如果由 $A_n\in\mathscr{M},n=1,2,\cdots,A_n\uparrow A$ 或 $A_n\downarrow A$, 可得 $A\in\mathscr{M}$.

设 $\mathscr{E}$ 是某一集系. 记 $\mu(\mathscr{E})$ 为包含 $\mathscr{E}$ 的最小单调类. (仿照引理 1 的证明, 同样可以证明最小单调类存在性.)

引理 2 代数 $\mathscr{A}$ 同时又是最小 $\sigma-$ 代数 $\sigma(\mathscr{A})$ 的充分必要条件是, $\mathscr{A}$ 是最小单调类.

证明 每一个 $\sigma-$ 代数显然是单调类. 现在设 $\mathscr{A}$ 是单调类, 而 $A_n\in\mathscr{A},n=1,2,\cdots$ 显然,

$$B_n=\bigcup_{i=1}^{n}A_i\in\mathscr{A}\text{ 且 }B_n\subseteq B_{n+1},n=1,2,\cdots.$$

从而根据单调类的定义,

$$B_n \uparrow \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}.$$

类似可得 $\bigcap_{i=1}^{\infty} A_i \in \mathscr{A}$. □

我们利用该引理证明, 说明如下 "$\sigma-$ 代数" 与 "单调类" 两概念之间联系的定理.

定理 1 设 $\mathscr{A}$ 是代数, 则

$$\mu(\mathscr{A}) = \sigma(\mathscr{A}). \tag{1}$$

证明 由引理 2, 知 $\mu(\mathscr{A}) \subseteq \sigma(\mathscr{A})$. 因此, 只需证明 $\mu(\mathscr{A})$ 是 $\sigma-$ 代数. 但是, 由于 $\mathscr{M} = \mu(\mathscr{A})$ 是单调类, 故仍由引理 2 知, 只需证明 $\mu(\mathscr{A})$ 是代数.

对于任意 $A \in \mathscr{M}$, 证明 $\overline{A} \in \mathscr{M}$. 为此运用如下以后常用的 "适当集合原理".

以

$$\widetilde{\mathscr{M}} = \{B : B \in \mathscr{M}, \overline{B} \in \mathscr{M}\}$$

表示 "使 $\mu(\mathscr{A})$ 是代数" 所有集合的全体. 显然 $\mathscr{A} \subseteq \widetilde{\mathscr{M}} \subseteq \mathscr{M}$. 现在证明 $\widetilde{\mathscr{M}}$ 是单调类.

设 $B_n \in \widetilde{\mathscr{M}}$, 则 $B_n \in \mathscr{M}, \overline{B}_n \in \mathscr{M}$, 因此

$$\lim \uparrow B_n \in \mathscr{M}, \lim \uparrow \overline{B}_n \in \mathscr{M}, \lim \downarrow B_n \in \mathscr{M}, \lim \downarrow \overline{B}_n \in \mathscr{M}.$$

从而

$$\overline{\lim \uparrow B_n} = \lim \downarrow \overline{B}_n \in \mathscr{M}, \overline{\lim \downarrow B_n} = \lim \uparrow \overline{B}_n \in \mathscr{M},$$
$$\overline{\lim \uparrow \overline{B}_n} = \lim \downarrow B_n \in \mathscr{M}, \overline{\lim \downarrow \overline{B}_n} = \lim \uparrow B_n \in \mathscr{M}.$$

即 $\widetilde{\mathscr{M}}$ 是单调类. 由于 $\widetilde{\mathscr{M}} \subseteq \mathscr{M}$, 可见 $\mathscr{M}$ 是最小单调类. 所以 $\widetilde{\mathscr{M}} = \mathscr{M}$, 并且如果 $A \in \mathscr{M} = \mu(A)$, 则 $\overline{A} \in \mathscr{M}$, 即类 $\mathscr{M}$ 关于补的运算封闭.

现在证明, 类 $\mathscr{M}$ 关于交的运算也封闭.

设 $A \in \mathscr{M}$, 且

$$\mathscr{M}_A = \{B : B \in \mathscr{M}, A \cap B \in \mathscr{M}\}$$

由等式

$$\lim \downarrow (A \cap B_n) = A \cap \lim \downarrow B_n, \lim \uparrow (A \cap B_n) = A \cap \lim \uparrow B_n,$$

可见 $\mathscr{M}_A$ 是单调类.

其次, 容易验证

$$(A \in \mathscr{M}_B) \Leftrightarrow (B \in \mathscr{M}_A). \tag{2}$$

现在设 $A \in \mathscr{A}$, 则由于 $\mathscr{A}$ 是代数, 可见对于任意 $B \in \mathscr{A}$, 集合 $A \cap B \in \mathscr{A}$, 因此

$$\mathscr{A} \subseteq \mathscr{M}_A \subseteq \mathscr{M}.$$

由于 $\mathscr{M}_A$ 是单调类, 而 $\mathscr{M}$ 是最小单调类, 可见对于任意 $A \in \mathscr{A}, \mathscr{M}_A = \mathscr{M}$. 由 (2) 式可见, 对于 $A \in \mathscr{A}, B \in \mathscr{M}$, 有

$$(A \in \mathscr{M}_B) \Leftrightarrow (B \in \mathscr{M}_A = \mathscr{M}).$$

从而, 如果 $A \in \mathscr{A}$, 则对于任意 $B \in \mathscr{M}$, 有

$$A \in \mathscr{M}_B.$$

因为 $A \in \mathscr{A}$ 是任意的, 所以由此得

$$\mathscr{A} \subseteq \mathscr{M}_B \subseteq \mathscr{M}.$$

因此, 对于任意 $B \in \mathscr{M}$,

$$\mathscr{M}_B = \mathscr{M},$$

即对于任意 $B \in \mathscr{M}, C \in \mathscr{M}$, 有 $C \bigcap B \in \mathscr{M}$.

从而, $\mathscr{M}$ 关于补和交 (从而关于并) 的运算是封闭类, 即 $\mathscr{M} = \mu(\mathscr{A})$ 是代数. 于是定理得证. □

由对以上所作证明的分析, 可见在考虑按 "适当集合原理" 形成的集系时, 这些集系关于某些理论上的集合运算的封闭性是重要的.

从这种观点出发, 在整个有关 "单调类" 的问题中, 引进所谓 "$\pi-$ 系" 和 "$\lambda-$ 系" 的概念是有益的. 其实, 在证明定理 1 时已经用到这些概念. 这些概念, 还可以用来证明涉及所研究问题的一系列命题 (例如, 定理 2), 而且往往比直接验证有关集系是 "单调类" 更加方便.

定义 2 ("$\pi - \lambda-$ 系") 设 Ω 是某个空间. 空间 Ω 子集的集系 $\mathscr{P}$, 称做 $\pi-$ 系, 如果它关于有限次的交运算封闭: 若 $A_1, \cdots, A_n \in \mathscr{P}$, 则 $\bigcap_{k=1}^n A_k \in \mathscr{P}, n \geqslant 1$.

空间 Ω 子集的集系 $\mathscr{L}$, 称做$\lambda-$ 系, 如果

$(\lambda_a) \Omega \in \mathscr{L}$,

$(\lambda_b)(A, B \in \mathscr{L}$ 且 $A \subseteq B) \Rightarrow (B \backslash A \in \mathscr{L})$,

$(\lambda_c)(A_n \in \mathscr{L}, n \geqslant 1,$ 且 $A_n \uparrow A) \Rightarrow (A \in \mathscr{L})$.

同时是 $\pi-$ 系和 $\lambda-$ 系的空间 Ω 中的子集的集系 $\mathscr{D}$, 称做邓肯 (E. B. Дынкин) $\pi - \lambda-$ 系或$d-$ 系.

注 2 指出如下事实是有益的: $\lambda-$ 系 的定义中的条件 (λ_a), (λ_b) 和 (λ_c), 等价于条件 (λ_a), (λ_b') 和 (λ_c'), 其中

(λ_b') 若 $A \in \mathscr{L}$, 则 $\overline{A} \in \mathscr{L}$,

(λ_c') 若 $A_n \in \mathscr{L}, n \geqslant 1, A_n \bigcap A_m = \varnothing (m \neq n)$, 则 $\bigcap A_n \in \mathscr{L}$.

还应指出, 显然任何代数都是 $\pi-$ 系.

设 $\mathscr{E}$ 是某一集系, 则以 $\pi(\mathscr{E}), \lambda(\mathscr{E})$ 和 $d(\mathscr{E})$ 分别表示包含 $\mathscr{E}$ 的 $\pi-$ 系, $\lambda-$ 系和 $d-$ 系.

由下面的定理, 可见 $\pi-\lambda-$ 系的作用. 为说明该定理的含义, 我们指出, 每一 $\sigma-$ 代数都是 $\lambda-$ 系. 然而, 逆命题一般不成立. 例如, 若 $\Omega = \{1,2,3,4\}$, 则集系

$$\mathscr{L} = \{\varnothing, \Omega, (1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$$

是 $\lambda-$ 系, 但不是 $\sigma-$ 代数.

不过, 若再补充要求 $\lambda-$ 系同时又是 $\pi-$ 系, 则所得 $\pi-\lambda-$ 系就是 $\sigma-$ 代数.

定理 2 (关于 $\pi-\lambda-$ 系) a) *任何 $\pi-\lambda-$ 系 $\mathscr{E}$ 都是 $\sigma-$ 代数.*

b) *设 $\mathscr{E}$ 是集合的 $\pi-$ 系, 则 $\lambda(\mathscr{E}) = d(\mathscr{E}) = \sigma(\mathscr{E})$.*

c) *设 $\mathscr{E}$ 是集合的 $\pi-$ 系, 而 $\mathscr{L}$ 是某一 $\lambda-$ 系且 $\mathscr{E} \subseteq \mathscr{L}$, 则 $\sigma(\mathscr{E}) \subseteq \mathscr{L}$.*

证明 a) (由于 (λ_a)) 系 $\mathscr{E}$ 包含 Ω, 且 (由于 (λ_b')) 关于求补及有限次交封闭 (根据 (λ_b') 及关于 $\mathscr{E}$ 是 $\pi-$ 系的假设), 而根据假设 $\mathscr{E}$ 是 $\pi-$ 系. 因此, (根据 §1 的定义 1) 集系 $\mathscr{E}$ 是代数. 为证明集系 $\mathscr{E}$ 也是 $\sigma-$ 代数, (根据 §1 的定义 4) 需要证明: 如果集合 $B_1, B_2, \cdots \in \mathscr{E}$, 则其并 $\bigcup_n B_n$ 也属于 $\mathscr{E}$.

设 $A_1 = B_1, A_n = B_n \bigcap \overline{A}_1 \bigcap \cdots \bigcap \overline{A}_{n-1}, n > 1$, 则根据 (λ_c'), $\bigcap A_n \in \mathscr{E}$. 由于 $\bigcap B_n = \bigcap A_n$, 则 $\bigcap B_n \in \mathscr{E}$.

于是,$\pi-\lambda-$ 系是 $\sigma-$ 代数.

b) 考虑 $\lambda-$ 系 $\lambda(\mathscr{E})$ 和 $\sigma-$ 代数 $\sigma(\mathscr{E})$. 如前面已经指出, 任何 $\sigma-$ 代数都是 $\lambda-$ 系. 那么, 由于 $\sigma(\mathscr{E}) \supseteq \mathscr{E}$, 则 $\sigma(\mathscr{E}) = \lambda(\sigma(\mathscr{E})) \supseteq \lambda(\mathscr{E})$. 因此, $\lambda(\mathscr{E}) \subseteq \sigma(\mathscr{E})$.

假如现在能证明, $\lambda-$ 系 $\lambda(\mathscr{E})$ 也是 $\pi-$ 系, 那么由命题 a), 可见 $\lambda(\mathscr{E})$ 是包含 $\mathscr{E}$ 的 $\sigma-$ 代数. 由于 $\sigma(\mathscr{E})$ 是包含 $\mathscr{E}$ 的*最小* $\sigma-$ 代数, 故由已经证明的 $\lambda(\mathscr{E}) \subseteq \sigma(\mathscr{E})$, 可见 $\lambda(\mathscr{E}) = \sigma(\mathscr{E})$.

因此, 只需要证明 $\lambda(\mathscr{E})$ 是 $\pi-$ 系.

仿照定理 1 的证明, 并且利用 "*适当集合原理*".

设

$$\mathscr{E}_1 = \{B \in \lambda(\mathscr{E}): \text{ 对于一切 } A \in \mathscr{E}, B \bigcap A \in \lambda(\mathscr{E})\}.$$

若 $B \in \mathscr{E}$, 则 $B \bigcap A \in \mathscr{E}$ (因为 $\mathscr{E}$ 是 $\pi-$ 系). 因而 $\mathscr{E} \subseteq \mathscr{E}_1$. 因为 (按 $\mathscr{E}_1$ 的定义) $\mathscr{E}_1$ 是 $\lambda-$ 系, 所以 $\lambda(\mathscr{E}) \subseteq \lambda(\mathscr{E}_1) = \mathscr{E}_1$. 另一方面, 按 $\mathscr{E}_1$ 的定义, 有包含关系 $\mathscr{E}_1 \subseteq \lambda(\mathscr{E})$.

从而 $\mathscr{E}_1 = \lambda(\mathscr{E})$.

现在设

$$\mathscr{E}_2 = \{B \in \lambda(\mathscr{E}): \text{ 对于一切 } A \in \lambda(\mathscr{E}), B \bigcap A \in \lambda(\mathscr{E})\}.$$

同 $\mathscr{E}_1$ 一样, $\mathscr{E}_2$ 也是 $\lambda-$ 系.

取集合 $B \in \mathscr{E}$, 那么根据 $\mathscr{E}_1$ 的定义, 对于 $A \in \mathscr{E}_1 = \lambda(\mathscr{E})$, 有 $B \bigcap A \in \lambda(\mathscr{E})$. 从而, 根据 $\mathscr{E}_2$ 的定义, 可见 $\mathscr{E} \subseteq \mathscr{E}_2$, 且 $\lambda(\mathscr{E}) \subseteq \lambda(\mathscr{E}_2) = \mathscr{E}_2$. 因此, $\lambda(\mathscr{E}) \supseteq \mathscr{E}_2$, 所以 $\lambda(\mathscr{E}) = \mathscr{E}_2$, 故对于一切 $A, B \in \lambda(\mathscr{E})$, 集合 $B \bigcap A \in \lambda(\mathscr{E})$, 即 $\lambda(\mathscr{E})$ 是 $\pi-$ 系. 这样, $\lambda(\mathscr{E})$ 是 $\pi-\lambda-$ 系 (故 $\lambda(\mathscr{E}) = d(\mathscr{E})$), 如前面所指出的那样, 由此可见 $\lambda(\mathscr{E}) = \sigma(\mathscr{E})$.

于是, 命题 b) 得证.

c) 由于 $\mathscr{E} \subseteq \mathscr{L}$ 且 $\mathscr{L}$ 是 $\lambda-$ 系, 可见 $\lambda(\mathscr{E}) \subseteq \lambda(\mathscr{L}) = \mathscr{L}$. 由 b) 知 $\lambda(\mathscr{E}) = \sigma(\mathscr{E})$. 于是, $\sigma(\mathscr{E}) \subseteq \mathscr{L}$. □

注 3　定理 2 的结果可以直接由定理 1 得到 (练习题 10).

下面两个命题的证明, 是 "适当集合原理" 和关于 $\pi-\lambda-$ 系的定理 2 之很好的演示.

引理 3　*设 $\mathbf{P}$ 和 $\mathbf{Q}$ 是可测空间 $(\Omega, \mathscr{F})$ 上的两个概率测度; $\mathscr{E}$ 是 $\mathscr{F}$ 中集合的 $\pi-$ 系, 且测度 $\mathbf{P}$ 和 $\mathbf{Q}$ 在 $\mathscr{E}$ 中的集合上相等. 那么, 这两个测度 $\mathbf{P}$ 和 $\mathbf{Q}$ 在 $\sigma-$ 代数 $\sigma(\mathscr{E})$ 上也相等. 特别, 若 $\mathscr{A}$ 是代数, 则测度 $\mathbf{P}$ 和 $\mathbf{Q}$ 在 $\sigma-$ 代数 $\sigma(\mathscr{A})$ 上也相等.*

证明　利用 "适当集合原理", 作为这样的集合, 取

$$\mathscr{L} = \{A \in \sigma(\mathscr{E}) : \mathbf{P}(A) = \mathbf{Q}(A)\}.$$

显然 $\Omega \in \mathscr{L}$; 如果 $A \in \mathscr{L}$, 则由于 $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A) = 1 - \mathbf{Q}(A) = \mathbf{Q}(\overline{A})$, 显然 $\overline{A} \in \mathscr{L}$. 如果 $A_1, A_2, \cdots \in \mathscr{L}$ 且两两不相交, 则由测度 $\mathbf{P}$ 和 $\mathbf{Q}$ 的可数可加性, 有

$$\mathbf{P}\left(\bigcup_n A_n\right) = \sum_n \mathbf{P}(A_n) = \sum_n \mathbf{Q}(A_n) = \mathbf{Q}\left(\bigcup_n A_n\right).$$

从而, 性质 (λ_a), (λ_b') 和 (λ_c') 成立, 因此 $\mathscr{L}$ 是 $\lambda-$ 系.

根据引理的条件 $\mathscr{E} \subseteq \mathscr{L}$, 故 $\mathscr{E}$ 是 $\pi-$ 系. 那么, 由定理 2 的命题 c), 可见 $\sigma(\mathscr{E}) \subseteq \mathscr{L}$. 由于适当集合的定义, 这一性质恰好表示测度 $\mathbf{P}$ 和 $\mathbf{Q}$ 在 $\sigma-$ 代数 $\sigma(\mathscr{E})$ 上相等. □

引理 4　*设 $\mathscr{A}_1, \mathscr{A}_2, \cdots, \mathscr{A}_n$ 是 (关于测度 $\mathbf{P}$) 相互独立的事件代数. 那么, $\sigma-$ 代数 $\sigma(\mathscr{A}_1), \sigma(\mathscr{A}_2), \cdots, \sigma(\mathscr{A}_n)$ (关于测度 $\mathbf{P}$) 也相互独立.*

证明　我们首先指出, 在概率的一般理论中, 集合以及集合系 (代数, $\sigma-$ 代数, $\cdots$) 的独立性的定义, 和初等概率论中独立性的定义完全一致 (见第一章 §3, 定义 $2 \sim 5$).

设 $A_2, \cdots, A_n$ 相应为 $\mathscr{A}_2, \cdots, \mathscr{A}_n$ 中的集合, 且

$$\mathscr{L}_1 = \left\{A \in \sigma(\mathscr{A}_1) : \mathbf{P}(A \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(A) \prod_{k=2}^{n} \mathbf{P}(A_k)\right\}. \tag{3}$$

现在证明 $\mathscr{L}_1$ 是 $\lambda-$ 系.

显然 $\Omega \in \mathscr{L}_1$, 即性质 (λ_a) 成立. 设 $A, B \in \mathscr{L}_1$, 而 $A \subseteq B$. 那么, 由

$$\mathbf{P}(A \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(A) \prod_{k=2}^{n} \mathbf{P}(A_k)$$

和

$$\mathbf{P}(B \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(B) \prod_{k=2}^{n} \mathbf{P}(A_k),$$

因此由第二式减去第一式, 得

$$\mathbf{P}((B \backslash A) \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(B \backslash A) \prod_{k=2}^{n} \mathbf{P}(A_k).$$

因而条件 (λ_b) 成立. 最后, 如果集合 $B_k \in \sigma(\mathscr{A}_1), k \geqslant 1$, 且 $B_k \uparrow B$, 则

$$B_k \cap A_2 \cap \cdots \cap A_n \uparrow B \cap A_2 \cap \cdots \cap A_n.$$

因此, 因为概率 $\mathbf{P}$ 的上连续性 (见 §1 的定理), 当 $k \to \infty$ 时对

$$\mathbf{P}(B_k \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(B_k) \prod_{i=2}^{n} \mathbf{P}(A_i)$$

求极限, 得

$$\mathbf{P}(B_k \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(B_k) \prod_{i=2}^{n} \mathbf{P}(A_i),$$

即条件 (λ_c) 成立.

因此, $\mathscr{L}_1$ 是 $\lambda-$ 系, 且 $\mathscr{L}_1 \supseteq \mathscr{A}_1$. 根据定理 2 的命题 c), 可得 $\mathscr{L}_1 \supseteq \sigma(\mathscr{A}_1)$.

于是, 证明了集系 $\sigma(\mathscr{A}_1), \mathscr{A}_2, \cdots, \mathscr{A}_n$ 相互独立.

对于集系 $\sigma(\mathscr{A}_2), \mathscr{A}_3, \cdots, \mathscr{A}_n, \sigma(\mathscr{A}_1)$ 进行类似的论证, 可见集系 $\sigma(\mathscr{A}_2), \mathscr{A}_3, \cdots, \mathscr{A}_n, \sigma(\mathscr{A}_1)$ 相互独立, 或等价于集系 $\mathscr{A}_3, \cdots, \mathscr{A}_n, \sigma(\mathscr{A}_1), \sigma(\mathscr{A}_2)$.

继续这一过程, 可得由 $\sigma-$ 代数 $\sigma(\mathscr{A}_1), \sigma(\mathscr{A}_2), \cdots, \sigma(\mathscr{A}_n)$ 构成的相互独立的集系. □

注 4 我们再次分析, 为使集系形成 $\sigma-$ 代数, 需要满足什么条件.

为此, 如果集系关于可数次交运算

$$A_1, A_2, \cdots \in \mathscr{E} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathscr{E}$$

封闭, 则称之为 π^*- 系.

那么, 由 $\sigma-$ 代数的定义可见, 如果某一代数 $\mathscr{E}$ 同时又是 π^*- 系, 则它就是 $\sigma-$ 代数.

基于 “$\pi-\lambda-$ 系” 概念的处理方法略有不同. 这里不是从 “代数” 的概念出发, 而是从 “$\lambda-$ 系” 概念出发. 并且, 如由定理 2 的命题 a) 可见, 若此 $\lambda-$ 系同时又是 $\pi-$ 系, 则它也是 $\sigma-$ 代数.

显然, 这恰好是上两种处理方法的区别所在.

当验证某集系是否$\sigma-$ 代数时, 我们从检验该集系是否代数开始, 就是说, 进行此检验只需考虑集合的有限和 (或交). “可数性” (这正式 “全部要做的”) 出现在下列情形: 需要检验一个集系是否 π^*- 系.

对于 “$\lambda-\pi-$ 方法”, 验证性质 “有关集系是 $\sigma-$ 代数”, 我们首先从 “确定该集系是否 $\lambda-$ 系” 开始, $\lambda-$ 系的性质 (λ_c) 或 (λ_c') 已经涉及“可数” 运算. 但是, 在第二阶段当验证该集系是否 $\pi-$ 系时, 我们只进行集合的有限交或和的运算.

在结束叙述有关 “单调类” 的有关结果时, 我们给出其一种函数形式. (第八章 §2 定理 1 之引理的证明, 可以作为引用的例子.)

定理 3 设 $\mathscr{E}$ 是 $\mathscr{F}$ 中集合的 $\pi-$ 系, 而 $\mathscr{H}$ 是实数值 $\mathscr{F}-$ 可测函数的全体, 具有下列性质:

(h_1) 若 $A\in\mathscr{E}$, 则函数 $I_A\in\mathscr{H}$;

(h_2) 若 $f\in\mathscr{H},h\in\mathscr{H}$, 则对于任何实数 $c, f+h\in\mathscr{H},cf\in\mathscr{H}$;

(h_3) 若函数 $h_n\in\mathscr{H},n\geqslant 1,0\leqslant h_n\uparrow h$, 则 $h\in\mathscr{H}$.

那么, 集系 $\mathscr{H}$ 包含一切关于 $\sigma-$ 代数 $\sigma(E)$ 可测的有界函数.

证明 设 $\mathscr{L}=\{A\in\mathscr{F}:I_A\in\mathscr{H}\}$. 由 (h_1), 可见 $\mathscr{E}\subseteq\mathscr{L}$. 由于 (h_2) 和 (h_3), 可见集系 $\mathscr{L}$ 是$\lambda-$ 系 (练习题 11). 因此由定理 2 的命题 c), 可见 $\sigma(\mathscr{E})\subseteq\mathscr{L}$. 从而, 由 $A\in\sigma(\mathscr{E})$, 可见 $I_A\in\mathscr{H}$. 根据性质 (h_2), 由此可见, 所有简单函数也属于集系 $\mathscr{H}$. (所谓简单函数, 即形如 $I_{A_i}, A_i\in\sigma(\mathscr{E})$ 的函数的有限线性组合). 最后, 由性质知, 一切关于 $\sigma-$ 代数 $\sigma(\mathscr{E})$ 可测的有界函数属于集系 $\mathscr{H}$. □

注 5 设 $X_1,\cdots,X_n$ 是可测空间 $(\Omega,\mathscr{F})$ 上的随机变量, $\mathscr{F}^X=\sigma(X_1,\cdots,X_n)$, 而 $f=f(\omega)$ 是 $\mathscr{F}^X-$ 可测函数. 那么, 存在博雷尔函数 $F=F(x_1,\cdots,x_n)$, 使 $f(\omega)=F(X_1(\omega),\cdots,X_n(\omega))$.

为证明这一命题, 只需利用定理 3. 在定理 3 中, 作为 “适当集合原理” 中函数的适当集合, 取非负博雷尔函数 $F=F(x_1,\cdots,x_n)$ 的集合 $\mathscr{H}$, 而作为集合 $\mathscr{E}$, 取

$$\mathscr{E}=\{\omega:X_1(\omega)\leqslant x_1,\cdots,X_n(\omega)\leqslant x_n;x_i\in\mathbb{R},i=1,\cdots,n\}.$$

由定理 3 知, 任意非负 $\mathscr{F}^X-$ 可测的函数 $f=f(\omega)$, 都可以通过非负博雷尔函数 $F=F(x_1,\cdots,x_n)$ 表示为 $f(\omega)=F(X_1(\omega),\cdots,X_n(\omega))$. 一般 (未必非负函数), 函数 f 通过极限过度可以表示为 $f=f^+-f^-$.

下面研究对于概率论最重要的各种可测空间 $(\Omega,\mathscr{F})$.

2. 可测空间 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 设 $\mathbb{R}=(-\infty,\infty)$ 是实数轴, 对于任意 $-\infty\leqslant a<b<\infty$, 设

$$(a,b]=\{x\in\mathbb{R}:a<x\leqslant b\}.$$

假定把区间 $(a,\infty]$ 理解为区间 (a,∞). (这样的假定之所以必要, 为使区间 $(-\infty,b]$ 的补集具有同样的形式, 即左开而右闭.)

以 $\mathscr{A}$ 表示集合 $\mathbb{R}$ 上, 有限个形如 $(a,b]$ 的不相交区间之和构成的集系:

$$A\in\mathscr{A},\quad 其中 A=\sum_{i=1}^{n}(a_i,b_i], n<\infty.$$

不难验证, 如果给 $\mathscr{A}$ 补充上空集 $\varnothing$, 则所得集系形成代数, 但不是 $\sigma-$ 代数, 因为如果 $A_n=(0,1-1/n]\in\mathscr{A}$, 但是

$$\bigcup_n A_n=(0,1)\notin\mathscr{A}.$$

设 $\mathscr{B}(\mathbb{R})$ 是包含集系 $\mathscr{A}$ 的最小 $\sigma-$ 代数 $\sigma(\mathscr{A})$. 在数学分析中, 这一 $\sigma-$ 代数有重要应用, 称做数轴上集合的**博雷尔代数**, 其中的元素称做**博雷尔集**.

若以 $\mathscr{I}$ 表示形如 $(a,b]$ 的区间 I 的集系, 而以 $\sigma(\mathscr{I})$ 表示包含 $\mathscr{I}$ 的最小 $\sigma-$ 代数, 则不难验证 $\sigma(\mathscr{I})$ 是博雷尔代数. 换句话说, 由于 $\sigma(\mathscr{I})=\sigma(\alpha(\mathscr{I}))$, 博雷尔代数, 可以不通过代数 $\mathscr{A}$, 由集系 $\mathscr{I}$ 得到.

由于对于 $a<b$,

$$(a,b)=\bigcup_{n=1}^{\infty}\left(a,b-\frac{1}{n}\right],\quad [a,b]=\bigcap_{n=1}^{\infty}\left(a-\frac{1}{n},b\right],$$

$$\{a\}=\bigcap_{n=1}^{\infty}\left(a-\frac{1}{n},a\right],$$

可见, 博雷尔代数, 除包含形如 $(a,b]$ 的区间外, 还包含单点集 $\{a\}$, 以及如下 6 种集合:

$$(a,b),[a,b],[a,b),(-\infty,b),(-\infty,b],(a,-\infty).\tag{4}$$

还应指出, 由 $\mathbb{R}$ 中的集系生成的一切最小 $\sigma-$ 代数, 都由 (4) 式中同一种不相交区间之有限和构成, 并且等于 $\mathscr{B}(\mathbb{R})$. 因此, 构造博雷尔代数 $\mathscr{B}(\mathbb{R})$, 可以不从形如 $(a,b]$ 的区间出发, 而从上述 6 种区间的任何一种出发.

有时, 不得不涉及扩展数轴上 $\overline{\mathbb{R}}=[-\infty,\infty]$ 的集合的代数 $\mathscr{B}(\overline{\mathbb{R}})$, 称做由 $\overline{\mathbb{R}}$ 上的集系诱导的、由形如

$$(a,b]=\{x\in\overline{\mathbb{R}}:a<x\leqslant b\},\quad -\infty\leqslant a<b\leqslant\infty$$

的不相交区间之有限和构成的最小 $\sigma-$ 代数, 其中 $(-\infty,b]$ 即集合 $\{x\in\overline{\mathbb{R}}:-\infty\leqslant x\leqslant b\}$.

注 1 对于可测空间 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$, 又常使用记号 $(\mathbb{R},\mathscr{B}),(\mathbb{R}^1,\mathscr{B}_1)$.

注 2 在数轴 $\mathbb{R}$ 上引进 (与通常的欧几里得度量 $|x-y|$ 等价的) 度量:

$$\rho_1(x,y)=\frac{|x-y|}{1+|x-y|},$$

而以 $\mathscr{B}_0(\mathbb{R})$ 表示, 由有限个 "形如 $S_\rho(x^0)$ 的不相交开集之和" 诱导的、最小 $\sigma-$ 代数, 其中对于 $\rho>0, x^0\in\mathbb{R}$,

$$S_\rho(x^0)=\{x\in\mathbb{R}:\rho_1(x,x^0)<\rho\}.$$

那么, $\mathscr{B}_0(\mathbb{R})=\mathscr{B}(\mathbb{R})$ (见练习题 7).

3. 可测空间 $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 设 $\mathbb{R}^n=\mathbb{R}\times\cdots\times\mathbb{R}$ 是 n 条 (份) 数轴的直积 (或欧几里得直积), 亦即有序数组 $x=(x_1,\cdots,x_n)$ 的集合, 其中 $-\infty<x_k<\infty, k=1,\cdots,n$. 集合 $I=I_1\times\cdots\times I_n$, 其中 $I_k=(a_k,b_k]$, 即集合 $\{x\in\mathbb{R}^n:x_k\in I_k, k=1,\cdots,n\}$ 称做*矩形*, 而 I_k 称做此矩形的边. 以 $\mathscr{I}$ 表示有限不相交矩形 I 之和构成的集合的全体. 由矩形系 $\mathscr{I}$ 生成的最小 $\sigma-$ 代数 $\sigma(\mathscr{I})$, 称做中集合的*博雷尔代数*, 记作 $\mathscr{B}(\mathbb{R}^n)$. 现在说明, 可以用另一种方式得到这一博雷尔代数.

与矩形 $I=I_1\times\cdots\times I_n$ 同时, 考虑具有*博雷尔边*的矩形 $B=B_1\times\cdots\times B_n$ (B_k 是数轴上的博雷尔集, 且在直积 $\mathbb{R}^n=\mathbb{R}\times\cdots\times\mathbb{R}$ 中占据第 k 个位置). 包含具有博雷尔边的一切矩形的最小 $\sigma-$ 代数, 记作

$$\mathscr{B}(\mathbb{R})\otimes\cdots\otimes\mathscr{B}(\mathbb{R}).$$

并且称做 $\sigma-$ 代数 $\mathscr{B}(\mathbb{R})$ 的*直积*. 我们证明, 实际上,

$$\mathscr{B}(\mathbb{R}^n)=\mathscr{B}(\mathbb{R})\otimes\cdots\otimes\mathscr{B}(\mathbb{R}).$$

换句话说, "由有限个不相交矩形 $B=B_1\times\cdots\times B_n$ 之和" 形成的集系生成的最小 $\sigma-$ 代数, 与 "具有博雷尔边的、有限个不相交矩形 $I=I_1\times\cdots\times I_n$ 之和" 形成的集系, 二者重合.

证明本质上依赖于下面的引理.

引理 5 *设 $\mathscr{E}$ 是 Ω 的某一集系, 集合 $B\subseteq\Omega$; 假设根据定义,*

$$\mathscr{E}\cap B=\{A\cap B:A\in\mathscr{E}\},\tag{5}$$

而 $\sigma(\mathscr{E}\cap B)$ 是由集系 $\mathscr{E}\cap B$ 生成的 B 的子集最小 $\sigma-$ 代数. 那么

$$\sigma(\mathscr{E}\cap B)=\sigma(\mathscr{E})\cap B.\tag{6}$$

证明 由于 $\mathscr{E}\subseteq\sigma(\mathscr{E})$, 可见

$$\mathscr{E}\cap B\subseteq\sigma(\mathscr{E})\cap B.\tag{7}$$

由于 $\sigma(\mathscr{E})\cap B$ 是 B 中的 $\sigma-$ 代数, 因此由 (7) 式, 可见

$$\sigma(\mathscr{E}\cap B)\subseteq\sigma(\mathscr{E})\cap B.$$

为证明相反的包含关系, 仍利用适当集合原理. 记

$$\mathscr{C}_B=\{A\in\sigma(\mathscr{E}):A\cap B\in\sigma(\mathscr{E}\cap B)\}.$$

由于 $\sigma(\mathscr{E})$ 和 $\sigma(\mathscr{E}\cap B)$ 是 $\sigma-$ 代数, 则 $\mathscr{C}_B$ 也是 $\sigma-$ 代数, 而且显然

$$\mathscr{E}\subseteq\mathscr{C}_B\subseteq\sigma(\mathscr{E}\cap B).$$

由 $\sigma(\mathscr{E})\subseteq\sigma(\mathscr{C}_B)=\mathscr{C}_B\subseteq\sigma(\mathscr{E})$, 可见 $\sigma(\mathscr{E})=\mathscr{C}_B$. 从而, 对于每一集合 $A\in\sigma(\mathscr{E})$, 有

$$A\cap B\subseteq\sigma(\mathscr{E}\cap B),$$

于是,$\sigma(\mathscr{E})\cap B\subseteq\sigma(\mathscr{E}\cap B)$. □

现在证明$\sigma-$ 代数 $\mathscr{B}(\mathbb{R}^n)$ 与 $\mathscr{B}(\mathbb{R})\otimes\cdots\otimes\mathscr{B}(\mathbb{R})$重合 (相等). 对于 $n=1$, 二者显然重合. 现在证明, 对于 $n=2$, 二者重合.

因为 $\mathscr{B}(\mathbb{R}^2)\subseteq\mathscr{B}(\mathbb{R})\otimes\mathscr{B}(\mathbb{R})$, 只需证明, 任意 $B_1\times B_2$ 属于 $\mathscr{B}(\mathbb{R}^2)$.

设 $\mathbb{R}^2=\mathbb{R}_1\times\mathbb{R}_2$, 其中 $\mathbb{R}_1$ 和 $\mathbb{R}_2$ 分别为 "第一条" 和 "第二条" 实数轴; $\widetilde{\mathscr{B}}_1=\mathscr{B}_1\times\mathbb{R}_2,\widetilde{\mathscr{B}}_2=\mathbb{R}_1\times\mathscr{B}_2$, 其中 $\mathscr{B}_1\times\mathbb{R}_2(\mathbb{R}_1\times\mathscr{B}_2)$ 是形如 $B_1\times\mathbb{R}_2(\mathbb{R}_1\times B_2)$ 集合的全体, 其中 $B_1\in\mathscr{B}_1(B_2\in\mathscr{B}_2)$. 假设 $\widetilde{\mathscr{I}}_1$ 和 $\widetilde{\mathscr{I}}_2$ 是 $\mathbb{R}_1$ 和 $\mathbb{R}_2$ 中的区间的全体, 而 $\widetilde{\mathscr{I}}_1=\widetilde{\mathscr{I}}_1\times\mathbb{R}_2,\widetilde{\mathscr{I}}_2=\mathbb{R}_1\times\widetilde{\mathscr{I}}_2$. 那么, 如果 $\widetilde{B}_1=B_1\times\mathbb{R}_2,\widetilde{B}_2=\mathbb{R}_1\times B_2$, 则由 (6) 式, 有

$$\begin{aligned}B_1\times B_2=\widetilde{B}_1\times\widetilde{B}_2\in\widetilde{\mathscr{B}}_1\cap\widetilde{B}_2&=\sigma(\widetilde{\mathscr{I}}_1)\cap\widetilde{B}_2\\&=\sigma(\widetilde{\mathscr{I}}_1\cap\widetilde{B}_2)\subseteq\sigma(\widetilde{\mathscr{I}}_1\cap\widetilde{\mathscr{I}}_2)=\sigma(\mathscr{I}_1\times\mathscr{I}_2),\end{aligned}$$

而这正是要证明的.

任意 $n>2$ 的情形用类似的方法证明. □

注 设 $\mathscr{B}_0(\mathbb{R}^n)$ 是由不相交开 "球" 集有限个 "形如 $S_\rho(x^0)$ 的之和" 诱导的最小 $\sigma-$ 代数, 其中对于 $\rho>0,x^0\in\mathbb{R}^n$,

$$S_\rho(x^0)=\{x\in\mathbb{R}^n:\rho_n(x,x^0)<\rho\},$$

而度量为

$$\rho_n(x,x^0)=\sum_{k=1}^{n}2^{-k}\rho_1(x_k,x_k^0),$$

其中 $x=(x_1,\cdots,x_n),x^0=(x_1^0,\cdots,x_n^0)$.

于是, $\mathscr{B}_0(\mathbb{R}^n)=\mathscr{B}(\mathbb{R}^n)$ (练习题 7).

4. 可测空间 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 空间 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 是建立具有无限步概率模型的基础, 因此它在概率论中有重要意义.

空间 $\mathbb{R}^\infty$ 是有序数列

$$x = (x_1, x_2, \cdots), \quad -\infty < x_k < \infty, k = 1, 2, \cdots$$

集合. 记 I_k 和 B_k 相应为区间 (a_k, b_k) 和 (坐标为 x_k 的) 第 k 条直线的博雷尔集. 考虑柱集:

$$\mathscr{I}(I_1 \times \cdots \times I_n) = \{x : x = (x_1, x_2, \cdots), x_1 \in I_1, \cdots, x_n \in I_n\}, \tag{8}$$

$$\mathscr{I}(B_1 \times \cdots \times B_n) = \{x : x = (x_1, x_2, \cdots), x_1 \in B_1, \cdots, x_n \in B_n\}, \tag{9}$$

$$\mathscr{I}(B^n) = \{x : x = (x_1, x_2, \cdots, x_n) \in B_n\}, \tag{10}$$

其中 B^n 是 $\mathscr{B}(\mathbb{R}^n)$ 中的博雷尔集. 柱集中的每一个 "柱" $\mathscr{I}(B_1 \times \cdots \times B_n)$ 或 $\mathscr{I}(B_n)$, 可以视为以 $\mathbb{R}^{n+1}, \mathbb{R}^{n+2}, \cdots$, 为底的柱, 因为

$$\mathscr{I}(B_1 \times \cdots \times B_n) = \mathscr{I}(B_1 \times \cdots \times B_n \times \mathbb{R}),$$
$$\mathscr{I}(B^n) = \mathscr{I}(B^{n+1}),$$

其中 $B^{n+1} = B^n \times \mathbb{R}$.

不相交柱集 $\mathscr{I}(I_1 \times \cdots \times I_n)$ 的有限和构成的集合是代数. 同样, 不相交柱集 $\mathscr{I}(B_1 \times \cdots \times B_n)$ 的并构成的集合也是代数. 柱集系 $\mathscr{I}(B^n)$ 也是代数. 记 $\mathscr{B}(\mathbb{R}^\infty)$, $\mathscr{B}_1(\mathbb{R}^\infty)$ 和 $\mathscr{B}_2(\mathbb{R}^\infty)$ 相应为包含 (8),(9) 和 (10) 式中一切集合的最小 $\sigma-$ 代数 ($\sigma-$代数 $\mathscr{B}_1(\mathbb{R}^\infty)$ 常表示为 $\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}) \otimes \cdots$). 显然

$$\mathscr{B}(\mathbb{R}^\infty) \subseteq \mathscr{B}_1(\mathbb{R}^\infty) \subseteq \mathscr{B}_2(\mathbb{R}^\infty).$$

实际上, 这三个 $\sigma-$ 代数重合.

为证明对于每一个 $n = 1, 2, \cdots$ 记

$$\mathscr{C}_n = \{A \subseteq \mathbb{R}^n : [x = (x_1, \cdots, x_n) \in A] \in \mathscr{B}(\mathbb{R}^\infty)\}.$$

设 $B^n \in \mathscr{B}(\mathbb{R}^n)$, 那么

$$B^n \in \mathscr{C}_n.$$

由于 $\mathscr{C}_n$ 是 $\sigma-$ 代数, 故

$$\mathscr{B}(\mathbb{R}^n) \subseteq \sigma(\mathscr{C}_n) = \mathscr{C}_n.$$

由此可见 $\mathscr{B}_2(\mathbb{R}^\infty) \subseteq \mathscr{B}(\mathbb{R}^\infty)$.

于是, $\mathscr{B}(\mathbb{R}^\infty) = \mathscr{B}_1(\mathbb{R}^\infty) = \mathscr{B}_2(\mathbb{R}^\infty)$.

以后, 我们称 $\mathscr{B}(\mathbb{R}^\infty)$ 中的集合为 $(\mathbb{R}^\infty)$ 中的博雷尔集.

注 设 $\mathscr{B}_0(\mathbb{R}^\infty)$ 是由不相交开 "球" 集有限个 "形如 $S_\rho(x^0)$ 的之和" 诱导的最小 $\sigma-$ 代数, 其中对于 $\rho > 0, x^0 \in \mathbb{R}^\infty$,

$$S_\rho(x^0) = \{x \in \mathbb{R}^\infty : \rho_\infty(x, x^0) < \rho\},$$

而度量为

$$\rho_\infty(x, x^0) = \sum_{k=1}^{\infty} 2^{-k} \rho_1(x_k, x_k^0),$$

其中 $x = (x_1, x_2, \cdots), x^0 = (x_1^0, x_2^0, \cdots)$. 于是, $\mathscr{B}(\mathbb{R}^\infty) = \mathscr{B}_0(\mathbb{R}^\infty)$ (练习题 7).

举几个 $\mathbb{R}^\infty$ 中的博雷尔集的例子.

(a) $\{x \in \mathbb{R}^\infty : \sup\ x_n > a\}, \{x \in \mathbb{R}^\infty : \inf\ x_n < a\}$;

(b) $\{x \in \mathbb{R}^\infty : \overline{\lim} x_n \leqslant a\}, \{x \in \mathbb{R}^\infty : \underline{\lim} x_n > a\}$, 其中

$$\overline{\lim} x_n = \inf_n \sup_{m \geqslant n}\ x_m, \quad \underline{\lim} x_n = \sup_n \inf_{m \geqslant n} x_m;$$

(c) $\{x \in \mathbb{R}^\infty : x_n \to x\}$ —— 使极限 $\lim x_n$ 存在并有限的 $x \in \mathbb{R}^\infty$ 的集合;

(d) $\{x \in \mathbb{R}^\infty : \lim x_n > a\}$;

(e) $\left\{x \in \mathbb{R}^\infty : \sum\limits_{n=1}^{\infty} |x_n| > a\right\}$;

(f) $\left\{x \in \mathbb{R}^\infty :\right.$ 至少对于一个 $\left. n \geqslant 1, \sum\limits_{k=1}^{n} x_k = 0\right\}$.

例如, 为验证 (a) 中的集合属于 $\mathscr{B}(\mathbb{R}^\infty)$, 只需注意到

$$\{x : \sup\ x_n > a\} = \bigcup_n \{x : x_n > a\} \in \mathscr{B}(\mathbb{R}^\infty),$$

$$\{x : \inf\ x_n < a\} = \bigcup_n \{x : x_n < a\} \in \mathscr{B}(\mathbb{R}^\infty).$$

5. 可测空间 $(\mathbb{R}^T, \mathscr{B}(\mathbb{R}^T))$ 其中 T 是任意集合, 而空间 $\mathbb{R}^T$ 是实函数 $x = (x_t), t \in T$, 的全体*). 我们主要考虑 T 是数轴上不可数子集的情形. 为简便和确定计, 现在可以假设 $T = [0, \infty)$.

考虑如下三种类型的柱集.

$$\mathscr{I}_{t_1, \cdots, t_n}(I_1 \times \cdots \times I_n) = \{x : x_{t_1} \in I_1, \cdots, x_{t_n} \in I_n\}, \tag{11}$$

$$\mathscr{I}_{t_1, \cdots, t_n}(B_1 \times \cdots \times B_n) = \{x : x_{t_1} \in B_1, \cdots, x_{t_n} \in B_n\}, \tag{12}$$

$$\mathscr{I}_{t_1, \cdots, t_n}(B^n) = \{x : (x_{t_1}, \cdots, x_{t_n}) \in B^n\}, \tag{13}$$

其中 I_k 是形如 $(a_k, b_k]$ 的集合, B_k 是数轴上的博雷尔集, 而 B^n 是 $\mathbb{R}^n$ 中的博雷尔集.

---

*) 对于 $\mathbb{R}^T$ 中的函数, 以后还使用下面的记号: $x = (x_t)_{t \in T}, x = (x_t), t \in T$.

集合 $\mathscr{I}_{t_1,\cdots,t_n}(I_1\times\cdots\times I_n)$ 是如下函数的全体: 在时刻 $t_1,\cdots,t_n$ 函数 "经过窗口" $I_1,\cdots,I_n$, 而在其余时刻取任意值 (图 24).

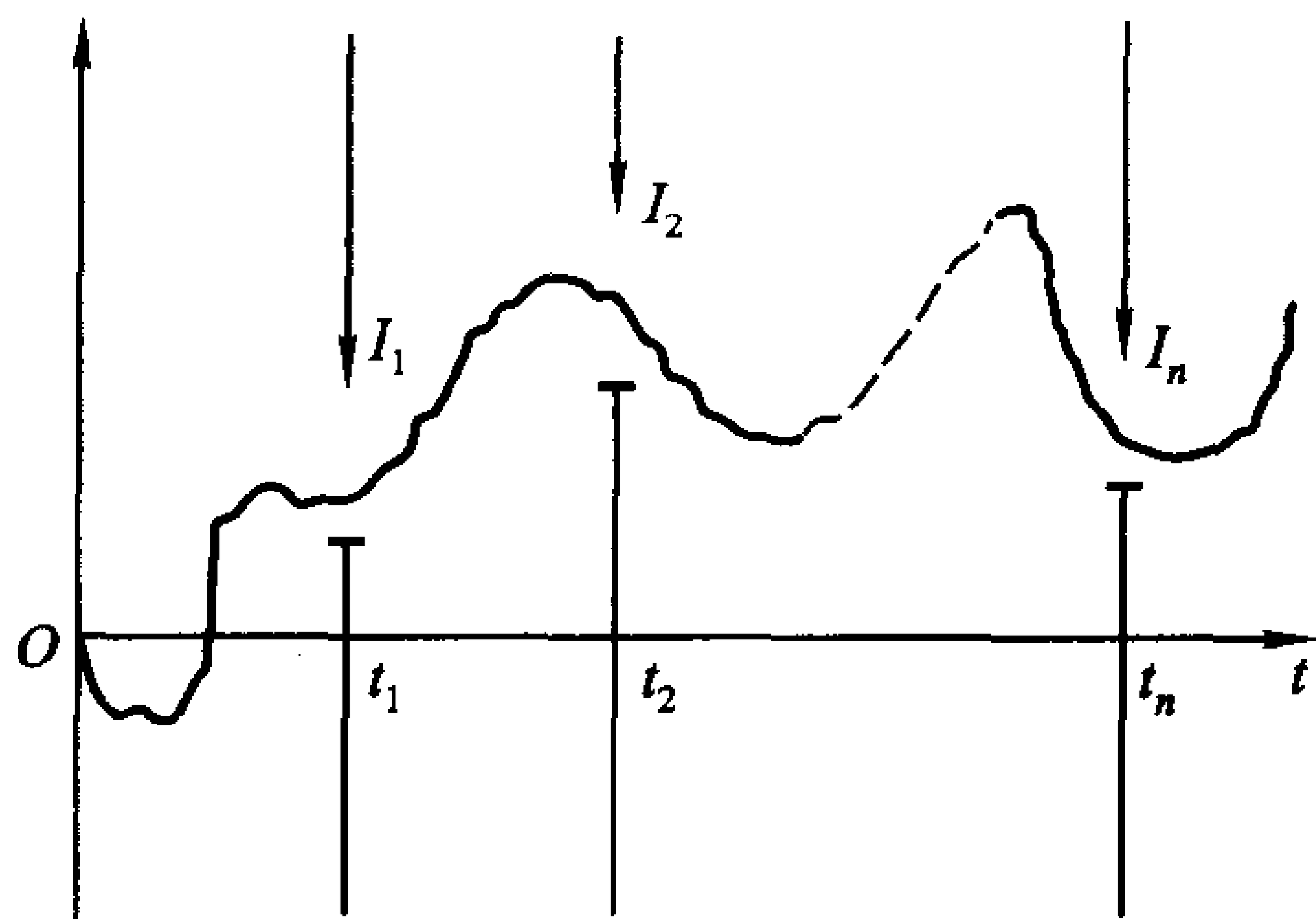


图 24

记 $\mathscr{B}(\mathbb{R}^T),\mathscr{B}_1(\mathbb{R}^T)$ 和 $\mathscr{B}_2(\mathbb{R}^T)$ 为相应地包含柱集 (11),(12) 和 (13) 式的最小 $\sigma-$ 代数. 显然

$$\mathscr{B}(\mathbb{R}^T)\subseteq\mathscr{B}_1(\mathbb{R}^T)\subseteq\mathscr{B}_2(\mathbb{R}^T). \tag{14}$$

然而, 实际上三个 $\sigma-$ 代数重合. 而且可以完全地描绘这些集合的结构.

定理 4 设 T 是不可数集合, 则 $\mathscr{B}(\mathbb{R}^T)=\mathscr{B}_1(\mathbb{R}^T)=\mathscr{B}_2(\mathbb{R}^T)$, 且任意集合 $A\in\mathscr{B}(\mathbb{R}^T)$ 有如下结构: 在 T 中存在最多可数个点 $t_1,t_2,\cdots$, 和博雷尔集 $B\in\mathscr{B}(\mathbb{R}^\infty)$, 使

$$A=\{x:x=(x_{t_1},x_{t_2},\cdots)\in B\}. \tag{15}$$

证明 设 $\mathscr{E}$ 是形如 (15) 式的集合的全体 (对于不同的数组 $(t_1,t_2,\cdots)$ 和 $B\in\mathscr{B}(\mathbb{R}^\infty)$). 若对应于 $A_1,A_2,\cdots\in\mathscr{E}$ 的数组为 $T^{(1)}=(t_1^{(1)},t_2^{(1)},\cdots),T^{(2)}=(t_1^{(2)},t_2^{(2)},\cdots)$, 则集合 $T^{(\infty)}=\bigcup_k T^{(k)}$ 可以取作一个统一集系: 一切 A_i 表示为

$$A_i=\{x:x=(x_{\tau_1},x_{\tau_2},\cdots)\in B_i\},$$

其中 B_i 是 (同一) $\sigma-$ 代数 $\mathscr{B}(\mathbb{R}^\infty)$ 的集合, 而 $\tau_i\in T^\infty$.

由此可见, 集系 $\mathscr{E}$ 是 $\sigma-$ 代数. 显然, 此 $\sigma-$ 代数包含一切形如 (13) 式的柱集, 而且由于 $\mathscr{B}_2(\mathbb{R}^T)$ 是包含这些集合的最小 $\sigma-$ 代数, 则由此并连同 (14) 式, 得

$$\mathscr{B}(\mathbb{R}^T)\subseteq\mathscr{B}_1(\mathbb{R}^T)\subseteq\mathscr{B}_2(\mathbb{R}^T)\subseteq\mathscr{E}. \tag{16}$$

考虑 $\mathscr{E}$ 中表示为 (15) 式的集合 A. 如果固定 $(t_1,t_2,\cdots)$, 则用与空间 $(\mathbb{R}^\infty,\mathscr{B}(\mathbb{R}^\infty))$ 情形同样的论述, 可见集合 A 是由 (11) 式的柱集生成的 $\sigma-$ 代数的元素. 然而, 这一 $\sigma-$ 代数显然属于 $\sigma-$ 代数 $\mathscr{B}(\mathbb{R}^T)$. 于是连同 (16) 式就可以证明定理 2 的两个命题. □

这样, $\sigma-$ 代数 $\mathscr{B}(\mathbb{R}^T)$ 中任何博雷尔集 A, 决定于 (至少在可数个点 $t_1, t_2, \cdots$ 上) 加在函数上 $x(x_t), t \in T$, 的约束. 特别, 由此可见, 依赖于函数在不可数点上的"性质" 的集合

$$A_1 = \{x: \text{ 对于一切 } t \in [0,1], x_t < C\},$$
$$A_2 = \{x: \text{ 至少对于一个 } t \in [0,1], x_t = 0\},$$
$$A_3 = \{x: \text{ 在固定的点 } t_0 \in [0,1], x_t \text{ 连续}\},$$

未必是博雷尔集. 而所提到的三个集合确实不属于$\mathscr{B}(\mathbb{R}^{[0,1]})$.

现在证明这对于 A_1 成立. 假如 $A_1 \in \mathscr{B}(\mathbb{R}^{[0,1]})$, 那么根据定理 4, 存在这样的点 $(t_1^0, t_2^0, \cdots)$ 和集合 $B^0 \in \mathscr{B}(\mathbb{R}^\infty)$, 使

$$\left\{x: \sup_t x_t < C, t \in [0,1]\right\} = \{x: x = (x_{t_1^0}, x_{t_2^0}, \cdots) \in B^0\}.$$

显然, 函数 $y_t \equiv C-1$ 属于 A_1, 从而 $(y_{t_1^0}, y_{t_2^0}, \cdots) \in B^0$. 定义函数

$$z_t = \begin{cases} C-1, & \text{若 } t \in (t_1^0, t_2^0, \cdots), \\ C+1, & \text{若 } t \notin (t_1^0, t_2^0, \cdots). \end{cases}$$

显然,

$$(y_{t_1^0}, y_{t_2^0}, \cdots) = (z_{t_1^0}, z_{t_2^0}, \cdots),$$

从而, 函数 $z = (z_t)$ 属于集合

$$\{x: (x_{t_1^0}, x_{t_2^0}, \cdots) \in B^0\}.$$

但是 $z = (z_t)$ 明显不属于集合 $\{x: \sup\ x_t < C\}$. 所得矛盾说明 $A_1 \notin \mathscr{B}(\mathbb{R}^{[0,1]})$.

由于关于一切函数 $x = (x_t), t \in [0,1]$ 空间中的 $\sigma-$ 代数 $\mathscr{B}(\mathbb{R}^{[0,1]})$, 集合 A_1, A_2 和 A_3 的不可测性, 自然地应考虑较窄的函数类, 以便使这些集合成为可测的. 直观上明显, 作为这样的空间, 例如可以考虑连续函数的空间.

6. 可测空间 $(C, \mathscr{B}(C))$ 设 $T = [0,1]$, 而 C 是连续函数 $x = (x_t), 0 \leqslant t \leqslant 1$ 的空间. 该空间关于均匀度量

$$\rho(x,y) = \sup_{t \in T} |x_t - y_t|$$

是可测空间. 在空间 C 中可以引进两个 $\sigma-$ 代数: $\mathscr{B}(C)$ 是由柱集生成的 $\sigma-$ 代数, 而 $\mathscr{B}_0(C)$ 是由 (关于度量 $\rho(x,y)$ 的)开集生成的 $\sigma-$ 代数. 现在证明, 两个 $\sigma-$ 代数实际上重合 $\mathscr{B}(C) = \mathscr{B}_0(C)$.

设 $B = \{x: x_{t_0} < b\}$ 是某一柱集. 易见, 此集合是开集. 由此可见

$$\{x: x_{t_1} < b_1, \cdots, x_{t_n} < b_n\} \in \mathscr{B}_0(C), \text{ 即 } \mathscr{B}(C) \subseteq \mathscr{B}_0(C).$$

相反, 考虑集合 $B_\rho = \{y : y \in S_\rho(x^0)\}$, 其中 x^0 是 C 中的某一函数, 而

$$S_\rho(x^0) = \{x \in C : \sup_{t\in T} |x_t - x_t^0| < \rho\}$$

是以 x^0 为球心的开球. 由 C 中函数的连续性, 可见

$$\begin{aligned} B_\rho &= \{y \in C : y \in S_\rho(x^0)\} = \{y \in C : \max_t |y_t - x_t^0| < \rho\} \\ &= \bigcap_{t_k} \{y \in C : |y_{t_k} - x_{t_k}^0| < \rho\} \in \mathscr{B}(C), \end{aligned} \tag{17}$$

其中 t_k 是区间 $[0,1]$ 上的有理点. 于是 $\mathscr{B}_0(C) \subseteq \mathscr{B}(C)$.

空间 $(C, \mathscr{B}_0(C), \rho)$ 是完备的而且是可分的; [5], [57].

7. 可测空间 $(D, \mathscr{B}(D))$　设 D 是一函数 $x = (x_t), 0 \leqslant t \leqslant 1$, 的空间, 其中每一个函数*右连续* $x_t = (x_{t+}), t < 1$, 而且对于任意 $t > 0$, 有*左极限*.

像空间 C 一样, 在空间 D 上可以定义度量 $d = d(x, y)$, 使 $\mathscr{B}_0(D) = \mathscr{B}(D)$, 其中 $\mathscr{B}_0(D)$ 是由*开集*生成的 $\sigma-$ 代数, $\mathscr{B}(D)$ 是由*柱集*生成的 $\sigma-$ 代数. 这时 $(D, \mathscr{B}(D), d)$ 是可分空间; [5], [57]. 关于*均匀度量*是斯科罗霍德 (A. B. Скороход) 引进的, 定义为:

$$d(x, y) = \inf\left\{\varepsilon > 0 : \exists \lambda \in \Lambda, \sup_t |x_t - y_{\lambda(t)}| + \sup_t |t - \lambda(t)| \leqslant \varepsilon\right\}, \tag{18}$$

其中 Λ 是 $[0,1]$ 上的严格递增连续函数 $\lambda(t)$, 且 $\lambda(0) = 0, \lambda(1) = 1$ 的集合.

8. 可测空间 $\left(\prod_{t\in T} \Omega_t, \boxtimes_{t\in T} \mathscr{F}_t\right)$　设空间 $(\mathbb{R}^T, \mathscr{B}(\mathbb{R}^T))$ 是 T "份" 数轴, 连同其博雷尔集系. 在概率论中, 除 $(\mathbb{R}^T, \mathscr{B}(\mathbb{R}^T))$ 外, 还考虑用如下方式构成的可测空间

$$\left(\prod_{t\in T} \Omega_t, \boxtimes_{t\in T} \mathscr{F}_t\right).$$

设 T 是下标的集合, $(\Omega_t, \mathscr{F}_t)$ 是可测空间 $t \in T$. 记

$$\Omega = \prod_{t\in T} \Omega_t$$

为一切函数 $\omega = (\omega_t), t \in T$, 的集合: 对于每个 $t \in T, \omega_t \in \Omega_t$.

不难验证, 一切有限个不相交柱集的并

$$\mathscr{I}_{t_1,\cdots,t_n}(B_1 \times \cdots \times B_n) = \{\omega : \omega_{t_1} \in B_1, \cdots, \omega_{t_n} \in B_n\}, B_{t_i} \in \mathscr{F}_{t_i},$$

构成代数. 以 $\boxtimes_{t\in T} \mathscr{F}_t$ 表示包含一切柱集的最小 $\sigma-$ 代数, 而可测空间 $\left(\prod \Omega_t, \boxtimes \mathscr{F}_t\right)$ 是可测空间 $(\Omega_t, \mathscr{F}_t)$ 的*直积*.

9. 练习题

1. 设 $\mathscr{B}_1$ 和 $\mathscr{B}_2$ 是空间 Ω 的子集的 $\sigma-$ 代数. 问集系

$$\mathscr{B}_1 \cap \mathscr{B}_2 \equiv \{A : A \in \mathscr{B}_1 \text{ 和 } A \in \mathscr{B}_2\},$$
$$\mathscr{B}_1 \cup \mathscr{B}_2 \equiv \{A : A \in \mathscr{B}_1 \text{ 或 } A \in \mathscr{B}_2\}$$

是否 $\sigma-$ 代数?

2. 设 $\mathscr{D} = \{D_1, D_2, \cdots\}$ 是 Ω 的某一可数分割, 而 $\mathscr{B} = \sigma(\mathscr{D})$. 问 $\sigma-$ 代数 $\mathscr{B}$ 的势如何?

3. 证明

$$\mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}) = \mathscr{B}(\mathbb{R}^{n+1}).$$

4. 证明第 4 小节的集合 (b)~(f) 属于 $\mathscr{B}(\mathbb{R}^\infty)$.

5. 证明第 5 小节的集合 A_2 和 A_3 不属于 $\mathscr{B}(\mathbb{R}^{[0,1]})$.

6. 证明函数 (18) 确实是度量.

7. 证明,$\mathscr{B}_0(\mathbb{R}^n) = \mathscr{B}(\mathbb{R}^n), n \geqslant 1$, 和 $\mathscr{B}_0(\mathbb{R}^\infty) = \mathscr{B}(\mathbb{R}^\infty)$.

8. 设 $C = C[0, \infty)$ 是定义连续函数 $x = (x_t), t \geqslant 0$ 的空间, 证明, 像 $C = C[0,1]$ 的情形一样, 该空间关于度量

$$\rho(x, y) = \sum_{n=1}^{\infty} 2^{-n} \min \left[\sup_{0 \leqslant t \leqslant n} |x_t - y_t|, 1 \right], x, y \in C,$$

是完备可分度量空间, 而 $\mathscr{B}_0(C) = \mathscr{B}(C)$, 其中 $\mathscr{B}_0(C)$ 是由开集生成的 $\sigma-$ 代数, $\mathscr{B}(C)$ 是由开集生成的 $\sigma-$ 代数.

9. 证明, (定义 2 的) 条件 $(\lambda_a), (\lambda_b), (\lambda_c)$, 与 (注 2 中的) 条件 $(\lambda_a), (\lambda_b'), (\lambda_c')$ 等价.

10. 由定理 1 证明定理 2.

11. 证明, 定理 3 中的 $\mathscr{L}$ 系是 $\lambda-$ 系.

12. 称 $\sigma-$ 代数为可数 – 生成的或可分的, 如果它有某个可数集类生成.

(a) 证明, 空间 $\Omega = (0, 1]$ 中博雷尔子集的 $\sigma-$ 代数 $\mathscr{B}$ 是可数 – 生成的.

(b) 证明, (例如) 如下的情形是可能的: 两个 $\sigma-$ 代数 $\mathscr{F}_1$ 和 $\mathscr{F}_2$, 且 $\mathscr{F}_1 \subset \mathscr{F}_2$, 但 $\mathscr{F}_2$ 可数生成, 而 $\mathscr{F}_1$ 不是可数生成的.

13. 证明, $\sigma-$ 代数 $\mathscr{G}$ 是可数 – 生成的充分和必要条件是: 对于某一随机变量 $X, \mathscr{G} = \sigma(X)$ ($\sigma(X)$ 的定义见 §4 第 4 小节).

14. 举例说明可分 $\sigma-$ 代数, 其相应的下 $-\sigma-$ 代数是不可分的.

15. 证明 $X_1, X_2, \cdots$ 是独立随机变量序列 §4, §5, 如果对于任意 $n \geqslant 1, \sigma(X_n)$ 和 $\sigma(X_1, \cdots, X_{n-1})$ 相互独立.

16. 举例说明两个 $\sigma-$ 代数之并, 不是 $\sigma-$ 代数.

17. 设 $\mathscr{A}_1$ 和 $\mathscr{A}_2$ 是两个独立的集系, 且每一个都是 $\pi-$ 系. 证明 $\sigma(\mathscr{A}_1)$ 和 $\sigma(\mathscr{A}_2)$ 也相互独立. 举例说明, 虽然两个不是 $\pi-$ 系的集系 $\mathscr{A}_1$ 和 $\mathscr{A}_2$ 独立, 但是 $\sigma(\mathscr{A}_1)$ 和 $\sigma(\mathscr{A}_2)$ 不独立.

18. 设 $\mathscr{L}$ 是 $\lambda-$ 系, 则

$$\{A, B \in \mathscr{L}, A \cap B = \varnothing\} \Rightarrow \{A \cup B \in \mathscr{L}\}.$$

19. 设 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 是在 Ω 上的子集的两个 $\sigma-$ 代数. 记

$$d(\mathscr{F}_1, \mathscr{F}_2) = 4 \sup_{\substack{A_1 \in \mathscr{F}_1 \\ A_2 \in \mathscr{F}_2}} |\mathbf{P}(A_1 A_2) - \mathbf{P}(A_1)\mathbf{P}(A_2)|.$$

证明, $d(\mathscr{F}_1, \mathscr{F}_2)$ 表征 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 的相依程度, 且具有如下性质:

(a) $0 \leqslant d(\mathscr{F}_1, \mathscr{F}_2) \leqslant 1$;

(b) 若 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 独立, 则 $d(\mathscr{F}_1, \mathscr{F}_2) = 0$;

(c) $d(\mathscr{F}_1, \mathscr{F}_2) = 1$, 当且仅当 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 的交含概率为 1/2 的事件.

20. 利用引理 1 的证明方法, 证明存在唯一含 $\mathscr{E}$ 集系 $\lambda-$ 系 $\lambda(\mathscr{E})$ 和 $\pi-$ 系 $\pi(\mathscr{E})$.

21. 设 $\mathscr{A}$ 是一集合代数: 对于任意不相交的集合序列 $(A_n)_{n \geqslant 1}, A_n \in \mathscr{A}$, 且

$$\bigcup_{n=1}^{\infty} A_n \in \mathscr{A},$$

证明 $\mathscr{A}$ 是 $\sigma-$ 代数.

22. 设 $(\mathscr{F}_n)_{n \geqslant 1}$ 是递增 $\sigma-$ 代数序列, $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}, n \geqslant 1$. 证明 $\bigcup_{n=1}^{\infty} \mathscr{F}_n$ 是 (一般刚好是) 代数.

23. 设 $\mathscr{F}$ 是代数 (或 $\sigma-$ 代数), 而 C 是 $\mathscr{F}$ 中某一集合. 考虑由 $\mathscr{F} \bigcup \{C\}$ 生成的最小代数 (相应地 $\sigma-$ 代数). 证明这一代数 (相应地 $\sigma-$ 代数) 的一切元素具有集合 $(A \bigcap C) \bigcup (B \bigcap \overline{C})$ 的形式, 其中 $A, B \in \mathscr{F}$.

24. 设 $\overline{\mathbb{R}} = \mathbb{R} \bigcup \{-\infty\} \bigcup \{\infty\}$ 是扩展的数轴. 博雷尔 $\sigma-$ 代数 $\mathscr{B}(\overline{\mathbb{R}})$ 定义为 (见第 2 小节): 由集合 $[-\infty, x], x \in \mathbb{R}$, 生成的 $\sigma-$ 代数, 其中 $[-\infty, x] = \{-\infty\} \bigcup (-\infty, x]$. 证明这一 $\sigma-$ 代数 $\mathscr{B}(\overline{\mathbb{R}})$ 与由集合生成的任何 $\sigma-$ 代数重合:

(a) $[-\infty, x), x \in \mathbb{R}$, 或

(b) $(x, \infty], x \in \mathbb{R}$, 或

(c) 一切有限区间 $\{-\infty\}$ 和与 $\{\infty\}$.

§3. 在可测空间上建立概率测度的方法

1. 可测空间 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$　设 $\mathbf{P}(A)$ 是定义在数轴的博雷尔集上的概率测度. 取 $A = (-\infty, x]$ 并设

$$F(x) = \mathbf{P}(-\infty, x], x \in \mathbb{R}. \tag{1}$$

这样定义的函数具有下列性质:

1) $F(x)$ 是非减函数;

2) $F(-\infty)=0, F(+\infty)=1$, 其中

$$F(-\infty)=\lim_{x\downarrow-\infty}F(x), F(+\infty)=\lim_{x\uparrow\infty}F(x);$$

3) $F(x)$ 在每一点 $x\in\mathbb{R}$ 右连续且有左极限.

第一条性质显然, 而另两条性质可以由概率测度的连续性得到.

定义 1 满足上述性质 1)~3) 的任意函数 $F(x)$, 称做 (数轴 $\mathbb{R}$ 上) 的**分布函数**.

于是, 与 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的每一个概率测度 $\mathbf{P}$, (由于 (1) 式) 有一个分布函数与之相对应. 逆命题仍然成立.

定理 1 *设 $F=F(x)$ 是数轴 $\mathbb{R}$ 上一分布函数. 那么, 在 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上存在唯一一个概率测度 $\mathbf{P}$, 使对于任意 $-\infty\leqslant a<b<\infty$, 有*

$$\mathbf{P}(a,b]=F(b)-F(a). \tag{2}$$

证明 设 $\mathscr{A}$ 是集合 $A\in\mathbb{R}$ 的代数, 其中每一个集合是形如 $(a,b]$ 的有限个不相交区间的和:

$$A=\sum_{k=1}^{n}(a_k,b_k].$$

在这些集合上定义一集函数 $\mathbf{P}_0$:

$$\mathbf{P}_0(A)=\sum_{k=1}^{n}[F(b_k)-F(a_k)], A\in\mathscr{A}. \tag{3}$$

此式在代数 $\mathscr{A}$ 上, 由 F 唯一确定一有限 – 可加集函数. 因此, 如果可以证明在代数 $\mathscr{A}$ 上函数 F 也是可数 – 可加的, 则概率测度 $\mathbf{P}$ 在 $\mathscr{B}(\mathbb{R})$ 上存在性和唯一性, 可以直接由测度论的一般结果得到 (这里不加证明而直接引用, 例如, 其证明可以参见 [42], [70]).

卡拉泰奥多里 (C. Carathéodory) **定理** *设 Ω 是某一空间, $\mathscr{A}$ 是其子集的代数, 而 $\mathscr{B}=\sigma(\mathscr{A})$ 是含 $\mathscr{A}$ 的最小 $\sigma-$ 代数. 记 μ_0 为 $(\Omega,\mathscr{A})$ 上的 $\sigma-$ 有限测度 (即 $\sigma-$ 可加集函数). 那么, 在 $(\Omega,\mathscr{B})$ 上存在且唯一测度 μ, 是 μ_0 的开拓, 即*

$$\mu(A)=\mu_0(A), A\in\mathscr{A}.$$

这样, 我们现在证明, 函数 $\mathbf{P}_0$ 在 $\mathscr{A}$ 上可数 – 可加 (即是概率测度). 根据 §1 的定理 1, 为此只需验证 $\mathbf{P}_0$ 在 $\varnothing$ 连续, 即

$$\mathbf{P}(A_n)\downarrow 0, A_n\downarrow\varnothing, A_n\in\mathscr{A}.$$

设 $A_1, A_2, \cdots$ 是自 $\mathscr{A}$ 选出的一集合序列, 且 $A_n \downarrow \varnothing$. 首先, 假设 A_n 属于某一闭区间 $[-N, N], N < \infty$. 因为 A_n 由形如 $(a, b]$ 的有限个区间的和构成, 且由于函数 $F(x)$ 的右连续性: 当 $a' \downarrow a$ 时

$$\mathbf{P}_0(a', b] = F(b) - F(a') \to F(b) - F(a) = \mathbf{P}_0(a, b],$$

可见对于每一个 A_n 存在一个集合 $B_n \in \mathscr{A}$, 使其闭包 $[B_n] \subseteq A_n$ 且

$$\mathbf{P}_0(A_n) - \mathbf{P}_0(B_n) \leqslant \varepsilon 2^{-n},$$

其中 ε 是某一事先给定的正数.

根据假设 $\bigcap A_n = \varnothing$, 因而 $\bigcap [B_n] = \varnothing$. 因为 $[B_n]$ 是闭集, 所以存在 $n_0 = n_0(\varepsilon)$, 使

$$\bigcap_{n=1}^{n_0} [B_n] = \varnothing. \tag{4}$$

实际上, $[-N, N]$ 是紧统, 而集系 $\{[-N, N] \backslash [B_n]\}_{n \geqslant 1}$ 是该紧统的开覆盖. 那么, 根据海涅 - 博雷尔 [H. E. Heine-E. Borel] 引理 (例如, 参见 [1],[33]), 存在有限子覆盖:

$$\bigcup_{n=1}^{n_0} ([-N, N] \backslash [B_n]) = [-N, N],$$

即

$$\bigcap_{n=1}^{n_0} [B_n] = \varnothing.$$

考虑到 (4) 式, 以及 $A_{n_0} \subseteq A_{n_0-1} \subseteq \cdots \subseteq A_1$, 有

$$\begin{aligned}\mathbf{P}_0(A_{n_0}) &= \mathbf{P}_0\left(A_{n_0} \backslash \bigcap_{k=1}^{n_0} B_k\right) + \mathbf{P}_0\left(\bigcap_{k=1}^{n_0} B_k\right) = \mathbf{P}_0\left(A_{n_0} \backslash \bigcap_{k=1}^{n_0} B_k\right) \\ &\leqslant \mathbf{P}_0\left(\bigcup_{k=1}^{n_0} (A_k \backslash B_k)\right) \leqslant \sum_{k=1}^{n_0} \mathbf{P}_0(A_k \backslash B_k) \leqslant \sum_{k=1}^{n_0} \varepsilon 2^{-k} \leqslant \varepsilon.\end{aligned}$$

因此 $\mathbf{P}_0(A_n) \downarrow 0, n \to \infty$.

现在去掉条件: 对于某个 $N, A_n \subseteq [-N, N]$. 对于给定的 $\varepsilon > 0$, 选择 N, 使 $\mathbf{P}_0[-N, N] > 1 - \varepsilon/2$. 那么, 由于

$$A_n = A_n \cap [-N, N] + A_n \cap \overline{[-N, N]},$$

可见

$$\begin{aligned}\mathbf{P}_0(A_n) &= \mathbf{P}_0(A_n \cap [-N, N]) + \mathbf{P}_0(A_n \cap \overline{[-N, N]}) \\ &\leqslant \mathbf{P}_0(A_n \cap [-N, N]) + \varepsilon/2,\end{aligned}$$

由上一段的结果 (只是将其中的 A_n 换成 $A_n \bigcap [-N,N]$) 可得, 对于充分大的 n, 有 $\mathbf{P}_0(A_n \bigcap [-N,N]) \leqslant \varepsilon/2$. 于是, 仍然得到 $\mathbf{P}_0(A_n) \downarrow 0, n \to \infty$. □

这样, 在 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的概率测度 $\mathbf{P}$ 与数轴上 $\mathbb{R}$ 的函数 F 之间, 存在一一对应关系. 习惯上, 把由函数 F 建立的测度 $\mathbf{P}$, 称做对应于分布函数 F 的**勒贝格 – 斯蒂尔切斯** (H. L. Lebesgue-T. J. Stieltjes) **概率测度**.

特别重要的情形是, 如果

$$F(x) = \begin{cases} 0, & 若\ x < 0, \\ x, & 若\ 0 \leqslant x \leqslant 1, \\ 1, & 若\ x > 1. \end{cases}$$

这时, 相应的概率测度 (记作 λ), 称做区间 $[0,1]$ 上的**勒贝格测度**. 显然, $\lambda(a,b] = b-a$, 其中 $0 \leqslant a \leqslant b \leqslant 1$. 换句话说, 区间 $(a,b]$ (以及区间 $(a,b), [a,b], [a,b)$) 上的勒贝格测度就等于区间的长度 $b-a$.

以

$$\mathscr{B}([0,1]) = \{A \cap [0,1] : A \in \mathscr{B}(\mathbb{R})\}$$

表示区间 $[0,1]$ 上**博雷尔**集合的全体. 除博雷尔集合外, 往往需要考虑区间 [0,1] 上所谓**勒贝格**集合. 称集合 $\Lambda \subseteq [0,1]$ 属于集系 $\overline{\mathscr{B}}([0,1])$, 如果存在博雷尔集合 A, B, 使 $A \subseteq \Lambda \subseteq B$ 且 $\lambda(B \backslash A) = 0$. 不难验证, $\overline{\mathscr{B}}([0,1])$ 是 $\sigma-$ 代数. 正是把 $\overline{\mathscr{B}}([0,1])$ 称做区间 [0,1] 上**勒贝格集合的集系**. 显然 $\mathscr{B}([0,1]) \subseteq \overline{\mathscr{B}}([0,1])$.

暂时仅定义在 $\mathscr{B}([0,1])$ 中集合上的测度 λ, 可以自然地开拓到**勒贝格集系** $\overline{\mathscr{B}}([0,1])$ 上. 具体地说, 如果 $\Lambda \in \overline{\mathscr{B}}([0,1])$ 且 $A \subseteq \Lambda \subseteq B$, 其中集系 $A, B \in \mathscr{B}([0,1])$ 且 $\lambda(B \backslash A) = 0$, 则设 $\overline{\lambda}(\Lambda) = \lambda(A)$. 不难验证, 这样定义的集函数 $\overline{\lambda} = \overline{\lambda}(\Lambda), \Lambda \in \overline{\mathscr{B}}([0,1])$, 是 $([0,1], \overline{\mathscr{B}}([0,1]))$ 上的概率测度, 称做 (勒贝格集系上的) **勒贝格测度**.

注 1 所采用的测度完备化 (开拓) 方法, 不只适用于所考虑的情形. 例如, 假设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是某一概率空间. 以 $\overline{\mathscr{F}}^{\mathbf{P}}$ 表示 Ω 的一切子集 Λ 的全体: 对于其中每一个子集 Λ, 存在 $A, B \in \mathscr{F}$, 使 $A \subseteq \Lambda \subseteq B$ 且 $\mathbf{P}(B \backslash A) = 0$. 自然, (利用等式 $\mathbf{P}(\Lambda) = \mathbf{P}(B)$) 也可以为 $\Lambda \in \overline{\mathscr{F}}^{\mathbf{P}}$ 定义概率测度. 用这种方法得到的概率空间 $(\Omega, \overline{\mathscr{F}}^{\mathbf{P}}, \mathbf{P})$, 称做空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 关于测度 $\mathbf{P}$ 的**完备化**.

如果对于概率测度 $\mathbf{P}$, 有 $\overline{\mathscr{F}}^{\mathbf{P}} = \mathscr{F}$, 则测度 $\mathbf{P}$ 称做**完备的**, 而相应的空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 称做**完备概率空间**.

注 2 现在简单地阐述卡拉泰奥多里定理的证明思路, 假设其中 $\mu_0(\Omega) = 1$.

设 A 是 Ω 中的集合, $A_1, A_2, \cdots$ 是 $\mathscr{A}$ 中的集合, 并且覆盖集合 $A : A \subseteq \bigcup_{n=1}^{\infty} A_n$. 定义集合 A 的**外测度** $\mu^*(A)$ 如下:

$$\mu^*(A) = \inf \sum_{n=1}^{\infty} \mu_0(A_n),$$

其中 inf 对集合 A 的一切上述覆盖 $(A_1, A_2, \cdots)$ 来求; 而

$$\mu_*(A) = 1 - \mu^*(\overline{A})$$

称做集合 A 的*内测度* $\mu_*(A)$.

设 $\widehat{\mathscr{A}}$ 是 Ω 中满足 $\mu_*(A) = \mu^*(A)$ 的集合 A 的全体. 不难证明, 集系 $\widehat{\mathscr{A}}$ 是 $\sigma-$代数 (练习题 12), 因而 $\mathscr{A} \subseteq \sigma(\mathscr{A}) \subseteq \widehat{\mathscr{A}}$. 赋予 $\widehat{\mathscr{A}}$ 中集合 A 以 "测度" $\mu(A)$, 令其等于 $\mu^*(A)(= \mu_*(A))$. 函数 $\mu(A)$ 确实是测度 (练习题 13), 即 $\mu(A)$ 确实是可数－可加集函数 (并且是概率测度, 因为 $\mu(\Omega) = \mu_0(\Omega) = 1$).

由等式 $\mathbf{P}(a, b] = F(b) - F(a)$ 建立的, 概率测度 $\mathbf{P}$ 与分布函数 F 之间的一一对应关系, 使得可以由相应的分布函数构造各种概率测度.

离散型测度　若分布函数 $F = F(x)$ 是*阶梯函数*, 则对应的测度称做离散型测度. 对于离散型测度, 分布函数仅在 (有限或可数个) 点 $x_1, x_2, \cdots$ 改变其数值: $\Delta F(x_i) > 0, i = 1, 2, \cdots$, 其中 $\Delta F(x) = F(x) - F(x-)$(图 25). 在这种情况下测度集中在点 $x_1, x_2, \cdots$ 上:

$$\mathbf{P}(\{x_k\}) = \Delta F(x_k) > 0, \quad \sum_k \mathbf{P}(\{x_k\}) = 1.$$

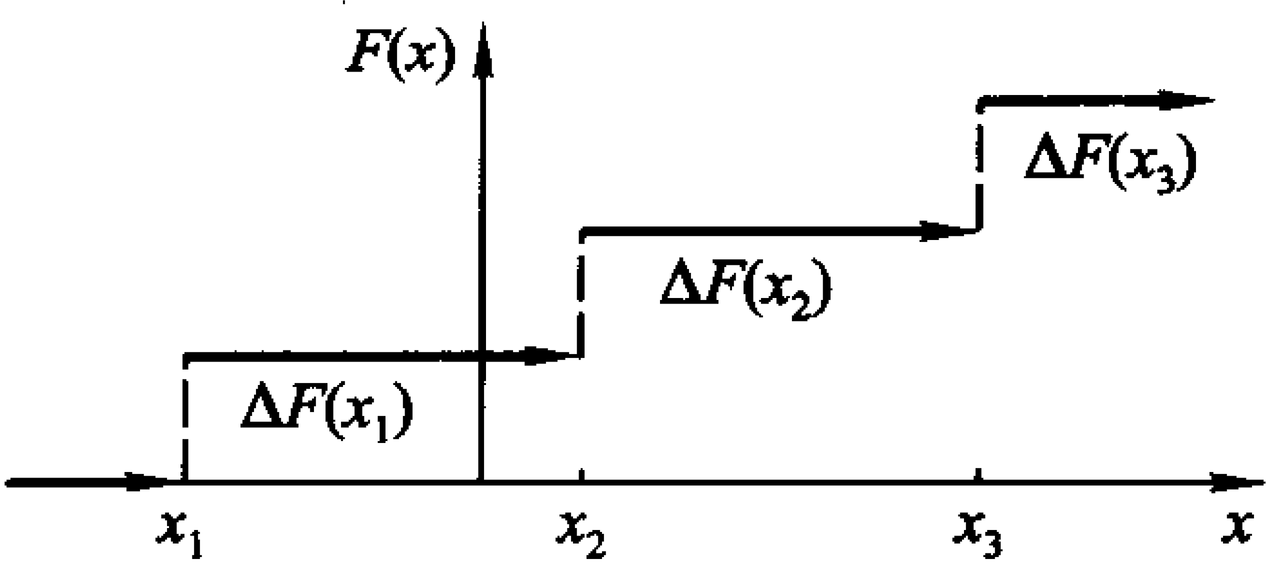


图 25

数组 $(p_1, p_2, \cdots)$, 其中 $p_k = \mathbf{P}(\{x_k\})$, 称做*离散型概率分布*, 而相应的分布函数 $F = F(x)$ 称做*离散型的*.

下表是最常见的离散概率分布类型及其名称.

表 2 — 2

分布名称	概　率　p_k	参　数
离散均匀	$\dfrac{1}{N}, k = 1, 2, \cdots, N$	$N = 1, 2, \cdots$
伯努利 (J. Bernoulli)	$p_1 = p, p_0 = q$	$0 \leqslant p \leqslant 1, q = 1 - p$
二　项	$\mathrm{C}_n^k p^k q^{n-k} (k = 0, 1, \cdots, n)$	$0 \leqslant p \leqslant 1, q = 1 - p (n = 1, 2, \cdots)$
泊　松 (S. D. Poisson)	$\dfrac{\lambda^k}{k!} \mathrm{e}^{-\lambda} (k = 0, 1, \cdots)$	$\lambda > 0$
几　何	$p q^{k-1} (k = 1, 2, \cdots)$	$0 < p \leqslant 1, q = 1 - p$
负二项 (帕斯卡) [B. Pascal]	$\mathrm{C}_{k-1}^{r-1} p^r q^{k-r} (k = r, r+1, \cdots)$	$0 < p \leqslant 1, q = 1 - p (r = 1, 2, \cdots)$

绝对连续测度 称测度为*绝对连续的*，如果存在非负博雷尔函数 $f(t), t \in \mathbb{R}$，使其分布函数 $F = F(x)$ 可以表示为:

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t, \tag{5}$$

其中的积分理解为黎曼 (G. F. B. Riemann) 积分 (而在一般情形下是勒贝格积分 (§6)).

函数 $f(t), t \in \mathbb{R}$，称做分布函数的*密度* (*概率分布密度*，或简称为*密度*)，而分布函数 $F = F(x)$ 本身称为*绝对连续的*.

表 2 − 3

分布名称	概率 p_k	参数
$[a,b]$ 上的均匀	$\frac{1}{b-a}(a \leqslant x \leqslant b)$	$a, b \in \mathbb{R}; a < b$
正态或高斯 (C. F. Gauss)	$\frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}, x \in \mathbb{R}$	$m \in \mathbb{R}, \sigma > 0$
伽 马 (Γ)	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \mathrm{e}^{-\frac{x}{\beta}}, x \geqslant 0$	$\alpha > 0, \beta > 0$
贝 塔 (B)	$\frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 \leqslant x \leqslant 1$	$\alpha > 0, \beta > 0$
指 数 (参数为 $\alpha = 1$, $\beta = \lambda^{-1}$ 的 Γ 分布)	$\lambda \mathrm{e}^{-\lambda x}, x \geqslant 0$	$\lambda > 0$
双侧指数	$\frac{\lambda}{2} \mathrm{e}^{-\lambda\|x-\alpha\|}, x \in \mathbb{R}$	$\lambda > 0, \alpha \in \mathbb{R}$
χ^2 (卡方) (Γ 分布 : $\alpha = 1/2, \beta = 2$)	$\frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}}, x \geqslant 0$	(自由度) $n = 1, 2, \cdots$
t (学生)	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, x \in \mathbb{R}$	(自由度) $n = 1, 2, \cdots$
F	$\frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\mathrm{B}\left(\frac{m}{2}, \frac{n}{2}\right)} x^{\frac{m}{2}-1} \left(1 + \frac{mx}{n}\right)^{-\frac{m+n}{2}}, x \geqslant 0$	(第一自由度) $m = 1, 2, \cdots$ (第二自由度) $n = 1, 2, \cdots$
柯 西 (A. L. Cauchy)	$\frac{\theta}{\pi(x^2 + \theta^2)}, x \in \mathbb{R}$	$\theta > 0$

显然，任何黎曼可积且在数轴上的积分为 1 即 $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ 的、非负博雷尔函数 $f = f(x), x \in \mathbb{R}$，由 (5) 式决定 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上某一概率测度的分布函数. 表 2–3

列出了概率论与数理统计中特别重要的, 不同类型的概率密度 $f = f(x)$ 的例子, 并且指出了其名称和参数. (在表中未指明 x 值的, 认为 $f(x) = 0$.)

奇异测度　称点 x 为分布函数 $F(x)$ 的增长点, 如果对于任意 $\varepsilon > 0$, 有 $F(x+\varepsilon) - F(x-\varepsilon) > 0$. 测度称为*奇异的*, 如果其分布函数 $F(x)$ 连续, *但是在其增长点集合的勒贝格测度*等于 0. 回避有关这样函数构造的细节 (例如, 可以参见 [70]), 我们仅限于举一个 "传统的" 例子.

用下面的康托尔 (G. Cantor) 方法, 构造区间 [0,1] 上的分布函数 $F(x)$.

将区间 [0,1] 分成三等份, 并且设 (图 26)

$$F_1(x) = \begin{cases} 1/2, & 若\ x \in (1/3, 2/3), \\ 0, & 若\ x = 0, \\ 1, & 若\ x = 1, \end{cases}$$

用线性内插法再给函数 $F_1(x)$ 补充上定义.

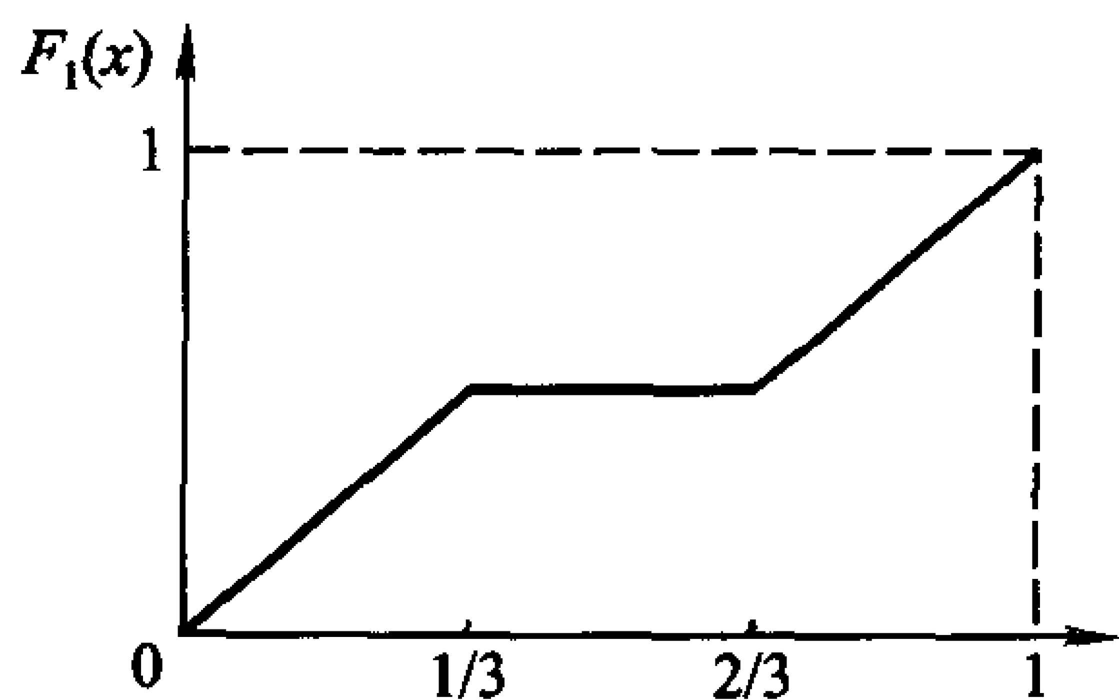


图 26

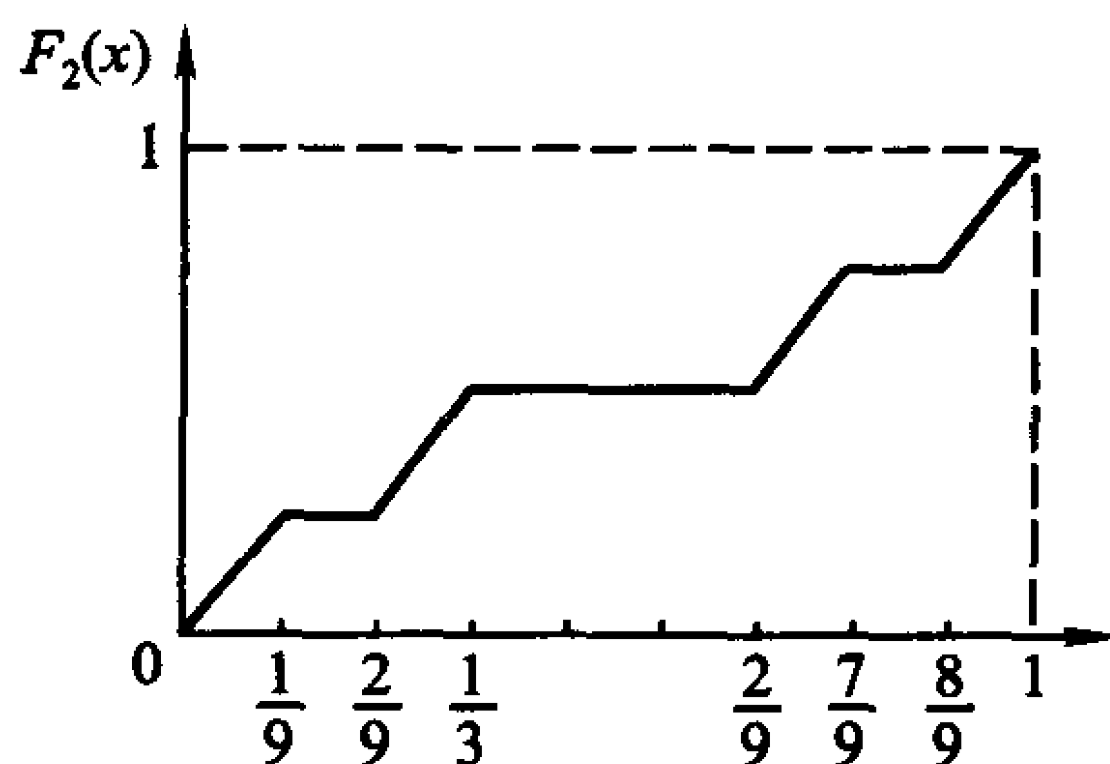


图 27

其次, 将区间 [0, 1/3] 和 [2/3, 1] 中每一个仍然分成三等份, 并且建立函数 (图 27):

$$F_2(x) = \begin{cases} 1/2, & 若\ x \in (1/3, 2/3), \\ 1/4, & 若\ x \in (1/9, 2/9), \\ 3/4, & 若\ x \in (7/9, 8/9), \\ 0, & 若\ x = 0, \\ 1, & 若\ x = 1, \end{cases}$$

也用线性内插法再给函数 $F_2(x)$ 补充上定义.

重复这一过程, 将建成函数序列 $F_n(x), n = 1, 2, \cdots$. 函数序列 $\{F_n(x)\}$ 收敛于某非减连续函数 $F(x)$ (称为康托尔函数), 而且其增长点的集合的勒贝格测度为 0. 事实上, 由函数 $F(x)$ 的构造过程可见, $F(x)$ 为常数的区间 $(1/3, 2/3), (1/9, 2/9), (7/9, 8/9), \cdots$ 的总长度为

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^n = 1. \tag{6}$$

以 $\mathscr{N}$ 表示康托尔 (G. Cantor) 函数 $F(x)$ 的增长点的集合. 由 (6) 式可见 $\lambda(\mathscr{N}) = 0$. 同时, 假如 μ 是对应于康托尔函数 $F(x)$ 的测度, 则 $\mu(\mathscr{N}) = 1$. (在这种情形下, 称测度 μ 关于勒贝格测度 λ 为*奇异的*.)

我们不准备过多的讨论关于分布函数的可能类型的问题, 只限于指出, 实际上上面指出的三种类型包含所有的分布函数. 确切地说, 任何分布函数 $F(x)$ 都可以表示为:

$$F(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x) + \alpha_3 F_3(x),$$

其中 $F_1(x)$ 是离散型分布函数, $F_2(x)$ 是连续型分布函数, $F_3(x)$ 是奇异型分布函数, 而 $\alpha_i (i = 1, 2, 3)$ 是非负实数, 而且 $\alpha_1 + \alpha_2 + \alpha_3 = 1$ (练习题 18).

2. 勒贝格测度到数轴上的开拓 定理 1 在 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的概率测度与 $\mathbb{R}$ 上的分布函数之间, 建立了一一对应关系. 分析该定理的证明可见, 实际上有更为一般的结果, 其中包括可以在整个数轴上引进所谓*勒贝格测度*.

设 μ 是 $(\Omega, \mathscr{A})$ 上某一 $\sigma-$ 有限测度, 其中 $\mathscr{A}$ 是 Ω 子集的代数. 结果表明, 关于测度 μ 自代数 $\mathscr{A}$ 开拓到最小 $\sigma-$ 代数 $\sigma(\mathscr{A})$ 的卡拉泰奥多里定理结论, 对于 $\sigma-$ 有限测度仍然成立, 这为定理 1 的推广提供了可能性.

使对于有限区间 I 的测度 $\mu(I) < \infty$ 的任何 $\sigma-$ 有限测度 μ, 称为 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的*勒贝格 - 斯蒂尔切斯测度*. 我们把定义在数轴 $\mathbb{R}$ 上, 值域为 $(-\infty, \infty)$ 的任何不减右连续函数 $G = G(x)$, 称为*广义分布函数*.

定理 1 可以这样推广, 使公式

$$\mu(a, b] = G(b) - G(a), a < b,$$

仍然可以建立勒贝格 - 斯蒂尔切斯测度 μ 与广义分布函数 G 之间的一一对应关系.

实际上, 如果 $G(+\infty) - G(-\infty) < \infty$, 则定理 1 的证明完全适用, 并且无需作任何改变, 因为可以将这种情形归结到 $G(+\infty) - G(-\infty) = 1$ 和 $G(-\infty) = 0$ 的情形.

现在假设 $G(+\infty) - G(-\infty) = \infty$. 设

$$G_n(x) = \begin{cases} G(x), & \text{若 } |x| \leqslant n, \\ G(n), & \text{若 } x > n, \\ G(-n), & \text{若 } x < -n. \end{cases}$$

在代数 $\mathscr{A}$ 上定义有限 - 可加测度 μ_0, 使其在 $(a, b]$ 的值 $\mu_0(a, b] = G(b) - G(a)$, 而设 μ_n 已经是 (按定理 1) 建立的、对应于函数的 $G_n(x)$ 的可数 - 可加测度.

显然, 在 $\mathscr{A}$ 上 $\mu_n \uparrow \mu_0$. 设 $A_1, A_2, \cdots$ 是 $\mathscr{A}$ 上的两两不相交集合, 且 $A \equiv \sum A_n \in \mathscr{A}$. 那么, (§1 练习题 6)

$$\mu_0(A) \geqslant \sum_{n=1}^{\infty} \mu_0(A_n).$$

而且, 如果 $\sum_{n=1}^{\infty}\mu_0(A_n)=\infty$, 则

$$\mu_0(A)=\sum_{n=1}^{\infty}\mu_0(A_n).$$

现在假设 $\sum\mu_0(A_n)<\infty$. 那么

$$\mu_0(A)=\lim_n\mu_n(A)=\lim_n\sum_{k=1}^{\infty}\mu_n(A_k).$$

根据所作的假设 $\sum\mu_0(A_n)<\infty$. 因此, 由 $\mu_n\leqslant\mu_0$ 可见

$$0\leqslant\mu_0(A)-\sum_{k=1}^{\infty}\mu_0(A_k)=\lim_n\left[\sum_{k=1}^{\infty}\mu_n(A_k)-\mu_0(A_k)\right]\leqslant 0.$$

这样, $\sigma-$ 有限可加测度 μ_0 在 $\mathscr{A}$ 上是有限 − 可加的, 因此 (根据卡拉泰奥多里定理) 它可以开拓到 $\sigma(\mathscr{A})$ 上的 $\sigma-$ 有限测度 μ.

$G(x)=x$ 的情形特别重要. 对应于这一广义分布函数的测度, 称为 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的*勒贝格测度*. 像区间 [0,1] 的情形一样, 在数轴上可以引进勒贝格集合系 $\overline{\mathscr{B}}(\mathbb{R})(\Lambda\in\overline{\mathscr{B}}(\mathbb{R})$, 如果存在勒贝格集合 A 和 B, 使 $A\subseteq\Lambda\subseteq B,\lambda(B\backslash A)=0)$, 对于 A 和 B 也可以定义勒贝格测度 $\overline{\lambda}$ (若 $A\subseteq\Lambda\subseteq B,\Lambda\in\overline{\mathscr{B}}(\mathbb{R})$ 且 $\lambda(B\backslash A)=0$, 则 $\overline{\lambda}(\Lambda)=\lambda(A)$).

3. 可测空间 $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 如同数轴的情形, 假设 $\mathbf{P}$ 是 $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 上的某一测度.

记

$$F_n(x_1,\cdots,x_n)=\mathbf{P}((-\infty,x_1]\times\cdots\times(-\infty,x_n]),$$

其更紧凑的形式为

$$F_n(x)=\mathbf{P}(-\infty,x],$$

其中 $x=(x_1,\cdots,x_n),(-\infty,x]=(-\infty,x_1]\times\cdots\times(-\infty,x_n]$.

引进 $\mathbb{R}^n\to\mathbb{R}$ 的差分算子 $\Delta_{a_ib_i}$, 按如下公式运作 $(a_i\leqslant b_i)$:

$$\begin{aligned}\Delta_{a_ib_i}F_n(x_1,\cdots,x_n)=&F_n(x_1,\cdots,x_{i-1},b_i,x_{i+1},\cdots,x_n)\\&-F_n(x_1,\cdots,x_{i-1},a_i,x_{i+1},\cdots,x_n).\end{aligned}$$

通过简单的运算, 可得

$$\Delta_{a_1b_1}\cdots\Delta_{a_nb_n}F_n(x_1,\cdots,x_n)=\mathbf{P}(a,b],\tag{7}$$

其中 $(a,b]=(a_1,b_1]\times\cdots\times(a_n,b_n]$. 特别, 一个不同于一维情形之处是, 一般

$$\mathbf{P}(a,b]\neq F_n(b)-F_n(a).$$

由于 $\mathbf{P}(a,b] \geqslant 0$. 故由 (7) 式可见, 对于任意 $a=(a_1,\cdots,a_n), b=(b_1,\cdots,b_n)$

$$\Delta_{a_1b_1}\cdots\Delta_{a_nb_n}F_n(x_1,\cdots,x_n) \geqslant 0. \tag{8}$$

由 $\mathbf{P}$ 的连续性, 可见 $F_n(x_1,\cdots,x_n)$ 对于变量的全体右连续, 即如果 $x=(x_1,\cdots,x_n), x^{(k)}=(x_1^{(k)},\cdots,x_n^{(k)})$, 则当 $x^{(k)}\downarrow x$ 时

$$F_n(x^{(k)})\downarrow F_n(x), k\to\infty. \tag{9}$$

同样明显,

$$F_n(+\infty,\cdots,+\infty)=1 \tag{10}$$

和

$$\lim_{x\downarrow y} F_n(x_1,\cdots,x_n)=0, \tag{11}$$

假如至少有一个 y 的坐标为 $-\infty$.

定义 2 满足条件 (8)~(11) 的任意函数 $F=F_n(x_1,\cdots,x_n)$, 称做 (空间 $\mathbb{R}^n$ 上的) *n 维分布函数*

运用与定理 1 同样的论述, 可以证明下面的定理.

定理 2 *设 $F=F_n(x_1,\cdots,x_n)$ 是 $\mathbb{R}^n$ 中某一分布函数, 则在 $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 上存在唯一概率测度 $\mathbf{P}$, 使*

$$\mathbf{P}(a,b]=\Delta_{a_1b_1}\cdots\Delta_{a_nb_n}F_n(x_1,\cdots,x_n). \tag{12}$$

举几个 n 维分布函数的例子.

设 $F^1,\cdots,F^n$ 是 $\mathbb{R}$ 上的一维分布函数, 而

$$F_n(x_1,\cdots,x_n)=F^1(x_1)\cdots F^n(x_n).$$

显然, 该函数右连续, 且满足条件 (10) 和 (11). 不难验证

$$\Delta_{a_1b_1}\cdots\Delta_{a_nb_n}F_n(x_1,\cdots,x_n)=\prod\left[F^k(b_k)-F^k(a_k)\right]\geqslant 0.$$

因而 $F_n(x_1,\cdots,x_n)$ 是一分布函数.

函数

$$F^k(x_k)=\begin{cases}0, & 若\ x_k<0,\\ x_k, & 若\ 0\leqslant x_k\leqslant 1,\\ 1, & 若\ x_k>1\end{cases}$$

的情形特别重要. 这时, 对于一切 $0\leqslant x_k\leqslant 1, k=1,\cdots,n$, 有

$$F_n(x_1,\cdots,x_n)=x_1\cdots x_n.$$

对应于这一 n 维分布函数的概率测度, 称做 $[0,1]^n$ 上的 n 维勒贝格测度.

多数 n 维分布函数具有如下形式

$$F_n(x_1,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}f_n(t_1,\cdots,t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n,$$

其中 $f_n(t_1,\cdots,t_n)$ 非负函数, 且满足

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_n(t_1,\cdots,t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n=1,$$

而积分可以理解为黎曼积分 (在更一般的情形下应理解为勒贝格积分). 函数 $f=f_n(t_1,\cdots,t_n)$ 称做 n维概率分布函数的密度, n 维概率分布密度, 或简称为 n 维密度.

当 $n=1$ 时, 函数

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}},x\in\mathbb{R},$$

其中 $\sigma>0$, 是 (非退化) 高斯分布密度或正态分布密度. 当 $n>1$ 时, 存在这一密度的自然类似情形.

设 $\boldsymbol{B}=(r_{ij})$ 是一 $n\times n$ 阶非负定对称矩阵:

$$\sum_{i,j=1}^{n}r_{ij}\lambda_i\lambda_j\geqslant 0,\lambda_i\in\mathbb{R},i=1,\cdots,n,$$
$$r_{ij}=r_{ji}.$$

当 $\boldsymbol{B}$ 是正定矩阵时, 其行列式 $|\boldsymbol{B}|\equiv\det\boldsymbol{B}>0$, 从而有逆矩阵 $\boldsymbol{A}=(a_{ij})\boldsymbol{B}^{-1}$. 那么, 函数

$$f_n(x_1,\cdots,x_n)=\frac{|\boldsymbol{A}|^{1/2}}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\sum a_{ij}(x_i-m_i)(x_j-m_j)\right\}, \tag{13}$$

其中 $m_i\in\mathbb{R},i=1,\cdots,n.f_n(x_1,\cdots,x_n)$ 具有下列性质: $f_n(x_1,\cdots,x_n)>0$, 且在整个空间上的 (n 重黎曼) 积分等于 1(这将在 §13 证明), 且由于它是正的, 故是概率密度.

这一函数称做 n 维 (非退化) 高斯分布密度或正态分布密度, 其均值向量为 $m=(m_1,\cdots,m_n)$, 而协方差矩阵为 $\boldsymbol{B}=\boldsymbol{A}^{-1}$.

当 $n=2$ 时, 密度 $f_2(x_2,x_2)$ 为:

$$\begin{aligned}f_2(x_2,x_2)=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\right.\\&\left.\times\left[\frac{(x_1-m_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}+\frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\},\end{aligned} \tag{14}$$

其中 $\sigma_i>0,|\rho|<1$. (关于参数 m_i,σ_i 和 ρ 的含义将在 §8 中说明.) 图 28 是二维正态分布的示意图.

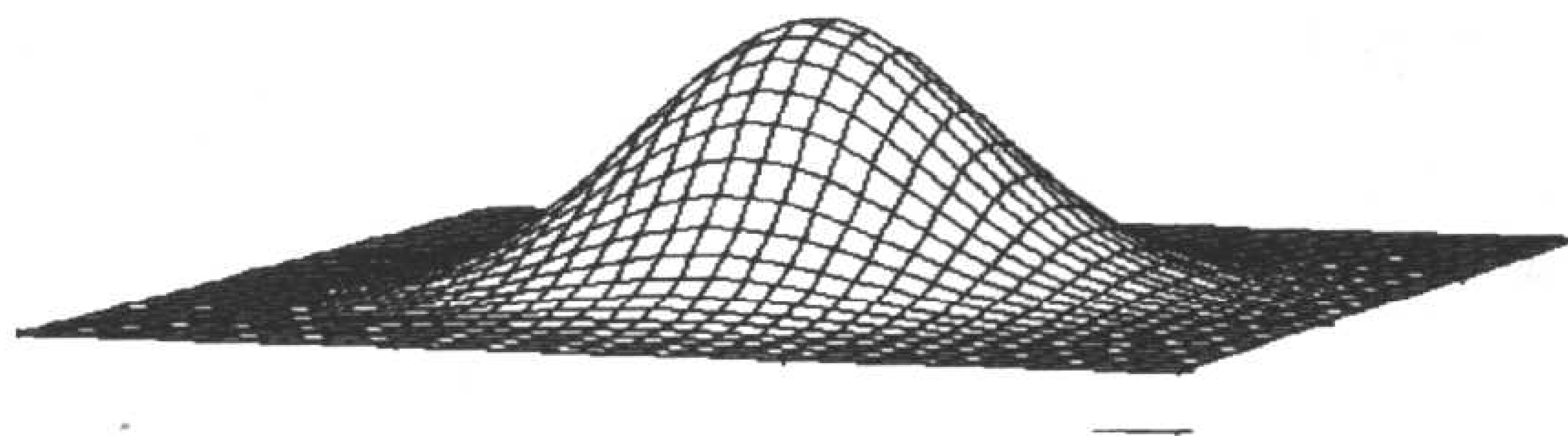


图 28 二维正态密度的图形

注 像 $n=1$ 的情形一样, 定理 2 可以推广到 (与定义类似的) $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ 上的勒贝格 – 斯蒂尔切斯测度, 以及 $\mathbb{R}^n$ 上的广义分布函数. 当广义分布函数 $G_n(x_1, \cdots, x_n)$ 等于 $x_1 \cdots x_n$ 时, 相应的测度称为空间 $\mathbb{R}^n$ 的博雷尔集合上的*勒贝格测度*. 显然, 对于博雷尔集合上的勒贝格测度,

$$\lambda(a, b] = \prod_{i=1}^{n}(b_i - a_i),$$

即 "矩形"

$$(a, b] = (a_1, b_1] \times \cdots \times (a_n, b_n]$$

的勒贝格测度等于其 "体积".

4. 可测空间 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 对于空间 $\mathbb{R}^n, n \geqslant 1$, 的情形, 概率测度是按如下模式建立的: 首先从基本集合 —— 形如 $(a, b]$ 的 "矩形" 出发, 然后将其自然地扩展到形如集合 $A = \sum(a_i, b_i]$, 最后根据卡拉泰奥多里定理将其开拓到 $\mathscr{B}(\mathbb{R}^n)$ 中的集合.

类似的建立概率模型的模式, 也 "适用于" 空间 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 的情形. 以

$$\mathscr{I}_n(B) = \{x \in \mathbb{R}^\infty : (x_1, \cdots, x_n) \in \mathscr{B}\}, B \in \mathscr{B}(\mathbb{R}^n)$$

表示空间 $\mathbb{R}^\infty$ 中 $B \in \mathscr{B}(\mathbb{R}^\infty)$ 为 "底" 的柱集的集合. 我们将看到, 正是把柱集自然地视为 $\mathbb{R}^\infty$ 中的*基本集合*, 并根据其概率的值定义 $\mathscr{B}(\mathbb{R}^\infty)$ 中集合上的概率测度.

假设 $\mathbf{P}$ 是空间 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 上的某一概率测度. 对于 $n = 1, 2, \cdots$, 记

$$P_n(B) = \mathbf{P}(\mathscr{I}_n(B)), B \in \mathscr{B}(\mathbb{R}^n). \tag{15}$$

相应地定义在 $(\mathbb{R}, \mathscr{B}(\mathbb{R})), (\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2)), \cdots$ 上的概率测度 $P_1, P_2, \cdots$ 序列, 具有如下明显的*一致性*: 对于 $n = 1, 2, \cdots$ 和 $B \in \mathscr{B}(\mathbb{R}^n)$,

$$P_{n+1}(B \times \mathbb{R}) = P_n(B). \tag{16}$$

特别值得注意的是, 相反的结果也成立.

定理 3 (柯尔莫戈洛夫关于 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 中的测度)　*假设 $P_1, P_2, \cdots$ 是 $(\mathbb{R}, \mathscr{B}(\mathbb{R})), (\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2)), \cdots$ 上, 具有一致性 (16) 的概率测度序列. 那么, 在 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 上存在唯一的概率测度 $\mathbf{P}$, 使对于每个 $n = 1, 2, \cdots$, 有*

$$\mathbf{P}(\mathscr{I}_n(B)) = P_n(B), B \in \mathscr{B}(\mathbb{R}^n). \tag{17}$$

证明　假设 $B \in \mathscr{B}(\mathbb{R}^n)$ 而 $\mathscr{I}_n(B^n)$ 是以 B^n 为 "底" 的柱集. 赋予该柱集以测度 $\mathbf{P}(\mathscr{I}_n(B^n)) = P_n(B^n)$.

现在证明, 由于一致性条件, 这样的定义是适定的, 即 $\mathbf{P}(\mathscr{I}_n(B^n))$ 的值与柱集的表现方法无关. 事实上, 假设同一柱集有两种表示方法:

$$\mathscr{I}_n(B^n) = \mathscr{I}_{n+k}(B^{n+k}).$$

那么, 由此可见, 若 $(x_1, \cdots, x_{n+k}) \in \mathbb{R}^{n+k}$, 则

$$(x_1, \cdots, x_n) \in B \Leftrightarrow (x_1, \cdots, x_{n+k}) \in B^{n+k}, \tag{18}$$

从而, 由 (16) 式, 有

$$\begin{aligned} P_n(B^n) &= P_{n+1}((x_1, \cdots, x_{n+1}) : (x_1, \cdots, x_n) \in B^n) = \cdots \\ &= P_{n+k}((x_1, \cdots, x_{n+k}) : (x_1, \cdots, x_n) \in B) = P_{n+k}(B^{n+k}). \end{aligned}$$

设 $\mathscr{A}(\mathbb{R}^\infty)$ 是一切柱集 $\widehat{B}^n = \mathscr{I}_n(B^n), B^n \in \mathscr{B}(\mathbb{R}^n), n = 1, 2, \cdots$ 的全体. 易见, $\mathscr{A}(\mathbb{R}^\infty)$ 是代数.

现在设 $\widehat{B}^1, \cdots, \widehat{B}^k$ 是 $\mathscr{A}(\mathbb{R}^\infty)$ 中的不相交集合. 不失普遍性, 可以假设: 对于某个 n, 有 $\widehat{B}_i = \mathscr{I}_n(B_i^n), i = 1, 2, \cdots, k$, 其中 $B_1^n, \cdots, B_k^n$ 是 $\mathscr{B}(\mathbb{R}^n)$ 中的两两不相交集合. 那么,

$$\mathbf{P}\left(\sum_{i=1}^k \widehat{B}_i\right) = \mathbf{P}\left(\sum_{i=1}^k \mathscr{I}_n(B_i^n)\right) = P_n\left(\sum_{i=1}^k \widehat{B}_i\right) = \sum_{i=1}^k P_n(\widehat{B}_i^n) = \sum_{i=1}^k \mathbf{P}(\widehat{B}_i),$$

即集函数 $\mathbf{P}$ 在代数 $\mathscr{A}(\mathbb{R}^\infty)$ 上有限 – 可加.

现在证明, $\mathbf{P}$ 在 "零" 连续 (从而在 $\mathscr{A}(\mathbb{R}^\infty)$ 上 $\sigma-$ 可加; 见 §1 的定理), 也就是说, 若当 $n \to \infty$ 时 $\widehat{B}_n \downarrow \varnothing$, 则当 $n \to \infty$ 时 $\mathbf{P}(\widehat{B}_n) \to 0$.

假设相反, 即设 $\mathbf{P}(\widehat{B}_n) \to \delta > 0$. 不失普遍性, 可以认为:

$$\widehat{B}_n = \{x : (x_1, \cdots, x_n) \in B_n\}, B_n \in \mathscr{B}(\mathbb{R}^n).$$

利用概率测度 P_n 在 $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ 中的如下性质 (见例 9): 如果 $B_n \in \mathscr{B}(\mathbb{R}^n)$, 则对于任意给定的 $\delta > 0$, 存在紧统 $A_n \in \mathscr{B}(\mathbb{R}^n)$, 使 $A_n \subseteq B_n, P_n(B_n \backslash A_n) \leqslant \delta/2^{n+1}$. 故若 $\widehat{A}_n = \{x : (x_1, \cdots, x_n) \in A_n\}$, 则

$$\mathbf{P}(\widehat{B}_n \backslash \widehat{A}_n) = P_n(B_n \backslash A_n) \leqslant \delta/2^{n+1}.$$

考虑集合 $\widehat{C}_n = \bigcap_{k=1}^n \widehat{A}_k$, 假设

$$\widehat{C}_n = \{x : x = (x_1, \cdots, x_n) \in C_n\}.$$

考虑到集合序列 $\widehat{B}_n$ 递降, 得

$$\mathbf{P}(\widehat{B}_n \backslash \widehat{C}_n) \leqslant \sum_{k=1}^n \mathbf{P}(\widehat{B}_n \backslash \widehat{A}_k) \leqslant \sum_{k=1}^n \mathbf{P}(B_k \backslash A_k) \leqslant \delta/2.$$

但是按假设的条件 $\mathbf{P}(\widehat{B}_n) \to \delta > 0$, 由此应得

$$\lim_n \mathbf{P}(\widehat{C}_n) \geqslant \frac{\delta}{2} > 0.$$

现在证明这与 $\widehat{C}_n \downarrow \varnothing$ 矛盾.

事实上, 在集合 $\widehat{C}_n$ 中各选一点 $\widehat{x}^{(n)} = (x_1^{(n)}, x_2^{(n)}, \cdots)$. 那么, 对于任意 $n \geqslant 1$, 有 $\widehat{x}^{(n)} = (x_1^{(n)}, \cdots, x_n^{(n)}) \in C_n$.

设 (n_1) 是序列 (n) 的这样一个子列, 使 $x_1^{(n_1)} \to x_1^0$, 其中 x_1^0 是 C_1 中某一点. (这样的子列存在, 因为一切 $x_1^{(n_1)} \in C_1$, 而 C_1 是紧统.) 由子列 (n_1) 选这样一个子列 (n_2), 使得 $(x_1^{(n_2)}, x_2^{(n_2)}) \to (x_1^0, x_2^0) \in C_2$. 同样, 设 $(x_1^{(n_k)}, \cdots, x_k^{(n_k)}) \to (x_1^0, \cdots, x_k^0) \in C_k$. 最后, 得对角序列 (m_k), 其中 m_k 是子列 (n_k) 的第 k 项. 那么, 对于任意 $i = 1, 2, \cdots$, 当 $m_k \to \infty$ 时, 有 $x_i^{(m_k)} \to x_i^0$, 并且对于任意 $n = 1, 2, \cdots$, 点 $(x_1^0, x_2^0, \cdots) \in \widehat{C}_n$. 显然, 这与假设 $\widehat{C}_n \downarrow \varnothing, n \to \infty$ 矛盾.

于是, 集函数 $\mathbf{P}$ 在代数 $\mathscr{A}(\mathbb{R}^\infty)$ 上 $\sigma-$ 可加, 说明根据卡拉泰奥多里定理, 可以将 $\mathbf{P}$ 开拓为 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 上的 (概率) 测度. □

注 现在考虑的情形, 空间 $\mathbb{R}^\infty$ 是 $\mathbb{R}$ 的*直积* $\mathbb{R}^\infty = \mathbb{R} \times \mathbb{R} \times \cdots$ 自然产生一个问题, 假如把 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 换成*可测空间* $(\Omega_i, \mathscr{F}_i)$ 的直积, 定理 3 是否仍然成立.

只要仔细分析一下该定理的证明就会发现, 用到的直线具有拓扑特点唯一性质, 就是本质上用到: "在 $\mathscr{B}(\mathbb{R}^n)$ 的任意集合 A 中存在*紧统*, 使其概率测度可以任意地接近集合 A 概率测度". 不过熟知, 不仅空间 $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ 具有该性质, 而且任何完备可分的、具有由开集生成的 $\sigma-$ 代数的可测空间都具有这一性质.

这样, 定理 3 对于如下情形仍然成立: $P_1, P_2, \cdots$ 是空间

$$(\Omega_1, \mathscr{F}_1), (\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2), \cdots$$

上的满足一致性的概率测度序列, 其中 $(\Omega_i, \mathscr{F}_i)$ 是完备可分度量空间, $\sigma-$ 代数 $\mathscr{F}_i$ 由开集生成, 而用 $(\Omega_1 \times \Omega_2 \times \cdots, \mathscr{F}_1 \otimes \mathscr{F}_2 \otimes \cdots)$ 代替 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$.

在 §9 (定理 2) 将证明, 对于*任意*可测空间 $(\Omega_n, \mathscr{F}_n)$, 如果测度 $P_n, n \geqslant 1$ 是以某种*特别方式*构造的情形, 定理 3 的结果仍然成立. 在 (关于所述可测空间, 或测度 $\{P_n\}$ 族结构的拓扑特点无任何假设的) 一般情形下, 定理 3 可能不正确. 下面的例子就说明这一点.

考虑空间 $\Omega=(0,1]$, 它显然不完备. 在此空间上按如下模式建立 $\sigma-$ 代数序列 $\mathscr{F}_1\subseteq\mathscr{F}_2\subseteq\cdots$. 对于一切 $n=1,2,\cdots$, 设

$$\varphi_n(\omega)=\begin{cases}1, \text{ 若 } 0<\omega<1/n,\\ 0, \text{ 若 } 1/n\leqslant\omega\leqslant 1,\end{cases}$$
$$\mathscr{C}_n=\{A\in\Omega: A=[\omega:\varphi_n(\omega)\in B], B\in\mathscr{B}(\mathbb{R})\}$$

而 $\mathscr{F}_n=\sigma\{\mathscr{C}_1,\cdots,\mathscr{C}_n\}$ 是包含集系 $\mathscr{C}_1,\cdots,\mathscr{C}_n$ 的最小 $\sigma-$ 代数. 显然 $F_1\subseteq F_2\subseteq\cdots$. 设 $\mathscr{F}=\sigma(\bigcup\mathscr{F}_n)$ 是包含集系一切 $\mathscr{F}_n$ 的最小 $\sigma-$ 代数. 考虑可测空间 $(\Omega,\mathscr{F}_n)$, 并在其上以如下方式定义概率测度 P_n:

$$P_n\{\omega:(\varphi_1(\omega),\cdots,\varphi_n(\omega))\in B^n\}=\begin{cases}1, \text{ 若 } (1,\cdots,1)\in B^n,\\ 0, \text{ 若不然},\end{cases}$$

其中 $B^n\in\mathscr{B}(\mathbb{R}^n)$. 不难验证, 测度族是一致的: 如果 $A\in\mathscr{F}_n$, 则 $P_{n+1}(A)=P_n(A)$. 然而可以证明, 在 $(\Omega,\mathscr{F})$ 上不存在概率测度 $\mathbf{P}$, 使其收缩 $\mathbf{P}|\mathscr{F}_n$(即将测度 $\mathbf{P}$ 仅局限在 $\mathscr{F}_n$ 中的集合上) 与 $P_n, n=1,2,\cdots$ 重合. 事实上, 假如这样的测度 $\mathbf{P}$ 存在. 那么, 对于任意 $n=1,2,\cdots$, 有

$$\mathbf{P}\{\omega:\varphi_1(\omega)=\cdots=\varphi_n(\omega)=1\}=P_n\{\omega:\varphi_1(\omega)=\cdots=\varphi_n(\omega)=1\}=1. \tag{19}$$

但是,

$$\{\omega:\varphi_1(\omega)=\cdots=\varphi_n(\omega)=1\}=(0,1/n)\downarrow\varnothing,$$

与 (19) 式矛盾, 从而也与函数 $\mathbf{P}$ 的可数可加性 (与在 “零” $\varnothing$ 的连续性) 矛盾.

现在举一个 $(\mathbb{R}^\infty,\mathscr{B}(\mathbb{R}^\infty))$ 上概率测度的例. 设 $F_1(x),F_2(x),\cdots$ 是一维分布函数序列. 定义函数 $G_1(x)=F_1(x), G_2(x_1,x_2)=F_1(x_1)F_2(x_2),\cdots$, 记 $P_1,P_2,\cdots$ 为与这些函数相对应空间 $(\mathbb{R},\mathscr{B}(\mathbb{R})),(\mathbb{R}^2,\mathscr{B}(\mathbb{R}^2)),\cdots$ 上的概率测度. 那么, 由定理 3 可见, 在 $(\mathbb{R}^\infty,\mathscr{B}(\mathbb{R}^\infty))$ 上存在一测度, 使

$$\mathbf{P}\{x\in\mathbb{R}^\infty:(x_1,\cdots,x_n)\in B\}=P_n(B), B\in\mathscr{B}(\mathbb{R}^n),$$

特别,

$$\mathbf{P}\{x\in\mathbb{R}^\infty:x_1\leqslant a_1,\cdots,x_n\leqslant a_n\}=F_1(a_1)\cdots F_n(a_n).$$

设 $F_i(x)$ 是伯努利分布函数:

$$F_i(x)=\begin{cases}0, \text{ 若 } x<0,\\ q, \text{ 若 } 0\leqslant x<1,\\ 1, \text{ 若 } x\geqslant 1.\end{cases}$$

设 Ω 是一切数值序列 $x=(x_1,x_2,\cdots),x_i=0,1,$ 的空间, $\mathscr{B}(\mathbb{R}^\infty)\bigcap\Omega$ 是 Ω 中博雷尔子集的 $\sigma-$ 代数. 可以证明, 在 $(\Omega,\mathscr{B}(\mathbb{R}^\infty)\bigcap\Omega)$ 上存在概率测度 $\mathbf{P}$, 使对于任意 $n=1,2,\cdots$, 有

$$\mathbf{P}\{x:x_1=a_1,\cdots,x_n=a_n\}=p^{\Sigma a_i}q^{n-\Sigma a_i}.$$

需要指出, 正是因为缺少这一结果, 我们在第一章 §5 中不能以 (8) 式的形式表述大数定律.

5. 可测空间 $(\mathbb{R}^T,\mathscr{B}(\mathbb{R}^T))$ 设 T 是下标 $t\in T$ 的任意集合, $\mathbb{R}_t$ 而是下标为 t 的数轴. 考虑不同下标 $t_i,t_i\in T$ 任意无序数组 $\tau=[t_1,\cdots,t_n],n\geqslant 1$; 设 P_τ 是 $(\mathbb{R}^\tau,\mathscr{B}(\mathbb{R}^\tau))$ 上的概率测度, 其中 $\mathbb{R}^\tau=\mathbb{R}_{t_1}\times\cdots\times\mathbb{R}_{t_n}$.

设 τ 在一切有限无序数组的集合取值, 则称概率测度族 $\{P_\tau\}$ 为一致的, 如果对于任意两个数组 $\tau=[t_1,\cdots,t_n]$ 和 $\sigma=[s_1,\cdots,s_k]$ 且 $\sigma\subseteq\tau$, 对于任何 $B\in\mathscr{B}(\mathbb{R}^\sigma)$, 有

$$\begin{aligned}&P_\sigma\{(x_{s_1},\cdots,x_{s_k}):(x_{s_1},\cdots,x_{s_k})\in B\}\\&=P_\tau\{(x_{t_1},\cdots,x_{t_n}):(x_{t_1},\cdots,x_{t_k})\in B\}.\end{aligned}\tag{20}$$

定理 4 ($(\mathbb{R}^T,\mathscr{B}(\mathbb{R}^T))$ 上的柯尔莫戈洛夫测度开拓定理) *设 $\{P_\tau\}$ 是 $(\mathbb{R}^T,\mathscr{B}(\mathbb{R}^T))$ 上的一致概率测度族, 则在 $(\mathbb{R}^T,\mathscr{B}(\mathbb{R}^T))$ 上存在唯一概率测度 $\mathbf{P}$, 使对于一切不同下标 $t_i\in T$ 的无序数组 $\tau=[t_1,\cdots,t_n],B\in\mathscr{B}(\mathbb{R}^\tau)$ 和 $\mathscr{I}_\tau=\{x\in\mathbb{R}^T:(x_{t_1},\cdots,x_{t_n})\in B\}$, 有*

$$\mathbf{P}\{\mathscr{I}_\tau(B)\}=P_\tau(B).\tag{21}$$

证明 设集合 $\widehat{B}\in\mathscr{B}(\mathbb{R}^T)$. 根据 §2 定理 3 存在有限或可数集合 $S=\{s_1,s_2,\cdots\}\subseteq T$, 使 $\widehat{B}=\{x:(x_{s_1},x_{s_2},\cdots)\in B\}$, 其中 $B\in\mathscr{B}(\mathbb{R}^S),\mathbb{R}^S=\mathbb{R}_{s_1}\times\mathbb{R}_{s_2}\times\cdots$. 换句话说, $\widehat{B}=\mathscr{I}_S(B)$ 是 "底" 为 $B\in\mathscr{B}(\mathbb{R}^S)$ 的柱集.

在这样的柱集 $\widehat{B}=\mathscr{I}_S(B)$ 上定义一集函数 $\mathbf{P}$:

$$\mathbf{P}(\mathscr{I}_S(B))=P_S(B),\tag{22}$$

其中由定理 3 知, 概率测度 P_S 存在.

我们现在证明, $\mathbf{P}$ 就是定理所断定存在的测度. 为此需要证明两点. 第一, 验证定义 (22) 适定, 即在用不同方式表示 $\widehat{B}$ 的情形下, $\mathbf{P}(\widehat{B})$ 的值不变; 第二, 集函数 $\mathbf{P}$ 可数可加.

这样, 设 $\widehat{B}=\mathscr{I}_{S_1}(B_1)$ 和 $\widehat{B}=\mathscr{I}_{S_2}(B_2)$. 那么, 显然 $\widehat{B}=\mathscr{I}_{S_1\bigcup S_2}(B_3)$, 其中 $B_3\in\mathscr{B}(\mathbb{R}^{S_1\bigcup S_2})$, 因此只需证明, 若 $S\subseteq S'$ 且 $B\in\mathscr{B}(\mathbb{R}^S)$, 则 $P_{S'}(B')=P_S(B)$, 其中

$$B'=\{(x_{s'_1},x_{s'_2},\cdots):(x_{s_1},x_{s_2},\cdots)\in B\},$$

而 $S'=\{s'_1,s'_2,\cdots\},S=\{s_1,s_2,\cdots\}$.

由于一致性条件 (20), 等式 (22) 可以直接由定理 3 得出, 从而证明 $\mathbf{P}(\widehat{B})$ 的值与集合 $\widehat{B}$ 的表示方法无关.

其次, 为验证集函数 $\mathbf{P}$ 的可数可加性, 假设 $\{\widehat{B}_n\}$ 是 $\mathscr{B}(\mathbb{R}^T)$ 中某一两两不交的集合序列, 则存在有限或可数集合 $S \subseteq T$, 使对于任何 $n \geqslant 1, \widehat{B}_n = \mathscr{I}_S(B_n)$, 其中 $B_n \in \mathscr{B}(\mathbb{R}^S)$. 由于 P_S 是概率测度, 则

$$\begin{aligned}\mathbf{P}\left(\sum \widehat{B}_n\right) &= \mathbf{P}\left(\sum \mathscr{I}_S(B_n)\right) = P_S\left(\sum B_n\right) \\ &= \sum P_S(B_n) = \sum \mathbf{P}\left(\mathscr{I}_S(B_n)\right) = \sum \mathbf{P}(\widehat{B}_n).\end{aligned}$$

最后, 直接由测度 $\mathbf{P}$ 的构造可得性质 (21). □

注 1 需要强调 T 是下标的任意集合. 这时由于定理 3 的注可见, 如果用任意完备可分度量空间 Ω_t (连同由开集生成的 $\sigma-$ 代数) 代替数轴 $\mathbb{R}_t$, 则定理仍然成立.

注 2 曾假定所讨论的概率测度族 $\{P_\tau\}$, 对于一切不同下标的*无序数组* $\tau = [t_1, \cdots, t_n]$ 是给定的. 对此应着重强调, 作为 $\tau = [t_1, \cdots, t_n]$ 的函数, 这些测度 P_τ 实质上是由不同点 $\{t_1\}, \cdots, \{t_n\}$ 构成的集合的函数. (例如, 数组 $[a, b]$ 和数组 $[b, a]$ 应视为同一数组, 因为二者都是由点 $\{a\}$ 和 $\{b\}$ 构成的同一集合.) 有时作为开始的取概率测度族 $\{P_\tau\}$, 其中 τ 在一切不同下标的*有序数组* $\tau = (t_1, \cdots, t_n)$ 的集合上取值. (这时, 数组 (a, b) 和数组 (b, a) 应视为不同的数组, 因为对于有序数组元素的先后顺序至关重要.) 在这种情形下, 为使定理 4 仍然成立, 除条件 (20) 外需要再增加一个*一致性条件*:

$$P_{(t_1,\cdots,t_n)}(A_{t_1} \times \cdots \times A_{t_n}) = P_{(t_{i_1},\cdots,t_{i_n})}(A_{t_{i_1}} \times \cdots \times A_{t_{i_n}}), \tag{23}$$

其中 $(i_1, \cdots, i_n)$ 是数 $(1, \cdots, n)$ 的任意排列, $A_{t_i} \in \mathscr{B}(\mathbb{R}_{t_i})$; 条件 (23) 作为概率测度 $\mathbf{P}$ 存在的必要条件, 由 (21) 式 (将 $P_{[t_1,\cdots,t_n]}(B)$ 换成 $P_{(t_1,\cdots,t_n)}(B)$) 可见是显然的.

我们下面总假设所考虑的数组 τ *是无序的*. 如果 T *是数轴*上的集合 (或为某一完全有序集合), 则不失普遍性, 可以认为对于所考虑的数组 $\tau = [t_1, \cdots, t_n]$, 有 $t_1 < t_2 < \cdots < t_n$. (例如, 集合 τ 由数值点 $\{a_1\}, \{a_2\}, \cdots, \{a_n\}$ 组成, 则 τ 总可以表示为 $\tau = [t_1, t_2, \cdots, t_n]$, 其中 $t_1 < t_2 < \cdots < t_n$.) 于是, 这是所有 "有限维" 概率只需给出这样的数组 $\tau = [t_1, t_2, \cdots, t_n]$, 其中 $t_1 < t_2 < \cdots < t_n$, 其中 $t_1 = \min(a_1, \cdots, a_n), t_n = \max(a_1, \cdots, a_n)$. 这样, 在这种情形下所有 "有限维" 概率, 只需对于这样的数组 $\tau = [t_1, t_2, \cdots, t_n]$ 给出就可以了, 其中 $t_1 < t_2 < \cdots < t_n$.

现在考虑 $T = [0, \infty)$ 的情形. 这种情形下,$\mathbb{R}^T$ 是一切实函数 $x = (x_t)_{t \geqslant 0}$ 的空间. $(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)}))$ 上概率测度的一个重要的例子是所谓*维纳* (N. Weiner) *测度*, 其构造如下.

考虑高斯概率密度族 $\{\varphi_t(y|x)\}_{t \geqslant 0}$ (是 y 的函数, 而 x 固定):

$$\varphi_t(y|x) = \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{-\frac{(y-x)^2}{2t}}, y \in \mathbb{R};$$

对于每一数组 $\tau = [t_1, t_2, \cdots, t_n], t_1 < t_2 < \cdots < t_n$, 集合 $B = I_1 \times \cdots \times I_n$, 其中 $I_k = (a_k, b_k)$, 以及由公式

$$\begin{aligned} P_\tau(B) &= P_\tau(I_1 \times \cdots \times I_n) \\ &= \int_{I_1} \cdots \int_{I_n} \varphi_{t_1}(a_1|0)\varphi_{t_2-t_1}(a_2|a_1)\cdots\varphi_{t_n-t_{n-1}}(a_n|a_{n-1})\mathrm{d}a_1\cdots\mathrm{d}a_n \end{aligned} \tag{24}$$

定义的测度 $P_\tau(B)$ (积分理解为黎曼积分). 然后, 对于每一柱集

$$\mathscr{I}_{t_1,\cdots,t_n}(I_1 \times \cdots \times I_n) = \{x \in \mathbb{R}^T : x_{t_1} \in I_1, \cdots, x_{t_n} \in I_n\}$$

定义集函数 $\mathbf{P}$, 设

$$\mathbf{P}\left(\mathscr{I}_{t_1,\cdots,t_n}(I_1 \times \cdots \times I_n)\right) = P_{[t_1,\cdots,t_n]}(I_1 \times \cdots \times I_n).$$

这样赋予柱集 $\mathscr{I}_{t_1,\cdots,t_n}(I_1 \times \cdots \times I_n)$ 测度的方法, 其直观含义如下.

集合 $\mathscr{I}_{t_1,\cdots,t_n}(I_1 \times \cdots \times I_n)$ 是于时间 $t_1, \cdots, t_n$ 通过 "窗口" $I_1, \cdots, I_n$ (§2 图 24) 的, 所有函数的集合. 我们把 $\varphi_{t_k-t_{k-1}}(a_k|a_{k-1})\mathrm{d}a_k$ 看作 "质点自点 a_{k-1} 出发, 经时间 $t_k - t_{k-1}$ 到达点 a_k 的 $\mathrm{d}a_k-$ 邻域" 的概率. 那么, 在 (24) 式中所考虑的密度的积, 表示运动的 "质点" 在时间区间 $[0, t_1], [t_2, t_3], \cdots, [t_{n-1}, t_n]$ 平移增量一定的独立性.

不难看到, 这样建立的测度族 $\{P_\tau\}$ 是一致的, 从而可以开拓为 $(\mathbb{R}^{[0,\infty]}, \mathscr{B}(\mathbb{R}^{[0,\infty]}))$ 上的测度. 这样得到的测度, 在概率论中起重要作用. 该测度是维纳引进的, 称为*维纳测度*.

6. 练习题

1. 设 $F(x) = \mathbf{P}(-\infty, x]$. 证明下列公式:

$$\begin{aligned} &\mathbf{P}(a,b] = F(b) - F(a), \mathbf{P}(a,b) = F(b-) - F(a), \\ &\mathbf{P}[a,b] = F(b) - F(a-), \mathbf{P}[a,b) = F(b-) - F(a-), \\ &\mathbf{P}(\{x\}) = F(x) - F(x-), \end{aligned}$$

其中 $F(x-) = \lim\limits_{y \uparrow x} F(y)$.

2. 证明公式 (7).

3. 证明定理 2.

4. 证明, 分布函数 $F = F(x)$ 在 $\mathbb{R}$ 上最多有可数个间断点. 问对于 $\mathbb{R}^n$ 上的分布函数有何相应的结果?

5. 考虑函数

$$G(x,y) = \begin{cases} 1, & \text{若 } x + y \geqslant 0, \\ 0, & \text{若 } x + y < 0, \end{cases}$$

$G(x,y) = [x + y]$ 是 $x + y$ 的整数部分.

证明, 两个函数都具有性质: 对于每一个自变量右连续、递增, 但不是 $\mathbb{R}^2$ 上的 (广义) 分布函数.

6. 设 μ 是对应于连续广义分布函数的勒贝格 – 斯蒂尔切斯测度. 证明, 若集合 A 最多是可数的, 则 $\mu(A)=0$.

7. 设 c 是连续统的势. 证明中博雷尔集 $\mathbb{R}^n$ 的势等于 c, 而勒贝格 σ– 代数的势等于 2^c.

8. 设 $(\Omega,\mathscr{F},\mathbf{P})$ 是某一概率空间, $\mathscr{A}$ 是 Ω 子集的代数, 而 $\sigma(\mathscr{A})=\mathscr{F}$. 利用适当集合原理, 证明对于任意 $\varepsilon>0$ 和 $B\in\mathscr{F}$, 存在集合 $A\in\mathscr{A}$, 使

$$\mathbf{P}(A\triangle B)\leqslant\varepsilon.$$

9. 设 $\mathbf{P}$ 是 $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 上的概率测度. 证明, 对于任意 $\varepsilon>0$ 和 $B\in\mathscr{B}(\mathbb{R}^n)$, 存在紧集 A_1 和开集 A_2, 使 $A_1\subseteq B\subseteq A_2$, 且 $\mathbf{P}(A_2\backslash A_1)\leqslant\varepsilon$. (该结果曾用于定理 3 的证明.)

10. 设 $\mathbf{P}$ 是给定概率测度. 验证由 $\mathbf{P}_\tau(B)=\mathbf{P}\{\mathscr{I}_\tau(B)\}$ 建立的测度族的一致性 (对照 (21) 式).

11. 验证表 2 – 2 和表 2 – 3 中的 "分布" 确实是概率分布.

12. 证明第 1 小节的注 2 中的 $\widehat{\mathscr{A}}$ 是 σ– 代数.

13. 证明第 1 小节的注 2 中的集函数 $\mu(A), A\in\widehat{\mathscr{A}}$, 是测度.

14. 举一例说明, 若测度 μ_0 是 $\mathscr{A}$ 上的有限 – 可加测度 (但不可数 – 可加), 则不可能开拓为 $\sigma(\mathscr{A})$ 上的可数 – 可加测度.

15. 证明, Ω 子集的 σ– 代数 $\mathscr{A}$ 上的任何有限 – 可加概率测度, 都可以开拓为 Ω 的一切子集上的有限 – 可加概率测度.

16. 设 $\mathbf{P}$ 是 Ω 子集的 σ– 代数 $\mathscr{F}$ 上给定的概率测度. 假设 $C\subseteq\Omega$, 但是 $C\notin\mathscr{F}$, 证明测度 $\mathbf{P}$ 可以开拓到 σ– 代数 $\sigma(\mathscr{F}\bigcup\{C\})$, 并且保持可数可加性.

17. 证明连续型分布函数 F 的承载子是一完全集合 (即 $\mathrm{supp}F$ 的承载子是闭集, 且具有如下性质: 若 $x\in\mathrm{supp}\ F,\varepsilon>0$, 则存在 $y\in\mathrm{supp}\ F$, 使 $0<|x-y|<\varepsilon$). 证明 (任意) 分布函数的承载子都是闭集.

18. 证明关于每一分布函数 F 结构的如下基本结果 (见第 1 小节末尾): 每一个分布函数都是凸组合

$$F=\alpha_1F_{\mathrm{d}}+\alpha_2F_{\mathrm{abc}}+\alpha_3F_{\mathrm{sc}},$$

其中 F_{d} 是离散型分布函数, F_{abc} 是绝对连续分布函数, F_{sc} 是奇异连续分布函数; 而 $\alpha_i\geqslant 0,\alpha_1+\alpha_2+\alpha_3=1$.

19. 证明, 对于康托尔函数和对于康托尔增长点的集合 $\mathscr{N}$ (与 supp F 的承载子重合) 中的每一个点 x, 有如下表现:

$$x=\sum_{k=1}^{\infty}\frac{\alpha_k(x)}{3k},$$

其中 $\alpha_k(x)=0$ 或 2, 而且对于这样的点, 有

$$F(x)=\sum_{k=1}^{\infty}\frac{\alpha_k(x)}{2^{k+1}}.$$

20. 设 C 是 $\mathbb{R}$ 上一闭集. 建立分布函数 F, 使其承载子为 $\operatorname{supp} F=C$.

21. 证明, 二项分布 (§2 第 1 小节) 的分布函数

$$B_n(m;p)=\mathbf{P}_n\{\nu\leqslant m\}=\sum_{k=0}^{m}\mathrm{C}_n^k p^k q^{n-k}$$

可以通过 (不完全) B 函数 $B_n(m;p)$ 表示:

$$B_n(m;p)=\frac{1}{\mathrm{B}(m+1,n-m)}\int_p^1 x^m(1-x)^{n-m-1}\mathrm{d}x.$$

22. 证明, 泊松分布的分布函数

$$F(n;\lambda)=\sum_{k=0}^{n}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

可以通过 (不完全) Γ 函数表示:

$$F(n;\lambda)=\frac{1}{n!}\int_\lambda^\infty x^n\mathrm{e}^{-x}\mathrm{d}x.$$

23. 在描绘分布密度 $f=f(x)$ 的形状时, 除均值和方差外, 标准特征是参数 "偏度" (skewness)

$$\alpha_3=\frac{\mu_3}{\sigma^3}$$

和参数 "峰度" (peakedness 或 kurtosis)

$$\alpha_4=\frac{\mu_4}{\sigma^4},$$

其中

$$\mu_k=\int_{-\infty}^{\infty}(x-\mu)^k f(x)\mathrm{d}x,\mu=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x,\quad \sigma^2=\mu_2.$$

对于表 2 – 3 中所引的分布, 讨论关于参数 α_3 和 α_4 的值的问题.

24. 对于服从参数 $\beta=1$ 的 Γ 分布的随机变量 X (见表 2 – 3), 证明

$$\mathbf{E}X^k=\frac{\Gamma(k+\alpha)}{\Gamma(\alpha)}.$$

特别, $\mathbf{E}X=\alpha,\mathbf{E}X^2=\alpha(\alpha+1)$, 因此 $\mathbf{D}X=\alpha$.

当 $\beta\neq 1$ 时求类似的公式.

25. 对于服从 B 分布的随机变量 X (见表 2 − 3), 证明

$$\mathbf{E}X^k = \frac{\mathrm{B}(r+k,s)}{\mathrm{B}(r,s)}.$$

26. 对于二项分布, 试验次数 n 固定, 考虑在 n 次试验中 "成功" 次数 ν 恰好等于 r 的概率 $\mathbf{P}_n\{\nu=r\}$. 此概率 $\mathbf{P}_n\{\nu=r\}=\mathrm{C}_n^r p^r q^{n-r}, 0\leqslant r\leqslant n$, 其中 p 是每次试验 "成功" 的概率. 这些概率形成二项分布 (n 给定). 如果考虑问题 "最早出现 r 次 '成功' (随机地) 发生在第 $\tau=k(k\geqslant r)$ 次试验" 的概率, 将导出负二项分布 (或逆二项分布). 证明, 事件 $\{\tau=k\}$ 的概率为

$$\mathbf{P}^r\{\tau=k\}=\mathrm{C}_{k-1}^{r-1}p^r q^{k-r},\quad k=r,r+1,\cdots,$$

其中 $r=1,2,\cdots$ (p 是每次试验 "成功" 的概率). 这些概率 $\mathbf{P}^r\{\tau=k\}, k=r,r+1,\cdots$, 的全体形成负二项分布. 证明, 对于给定的 r, $\mathbf{E}^r\tau=rq/p$.

§4. 随机变量 I

1. 随机变量及其概率分布和分布函数　设 $(\Omega,\mathscr{F})$ 是某一可测空间, $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 是数轴及其博雷尔集系 $\mathscr{B}(\mathbb{R})$ 的可测空间.

定义 1　定义在 $(\Omega,\mathscr{F})$ 上的实函数 $\xi=\xi(\omega)$, 称做 $\mathscr{F}$ − 可测函数或随机变量, 如果对于任意 $B\in\mathscr{B}(\mathbb{R})$, 有

$$\{\omega:\xi(\omega)\in B\}\in\mathscr{F}; \tag{1}$$

或者, 如果 $\xi^{-1}(B)\equiv\{\omega:\xi(\omega)\in B\}$ 是 Ω 中的可测集.

当 $(\Omega,\mathscr{F})=(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 时, $\mathscr{B}(\mathbb{R}^n)$− 可测函数称做博雷尔可测函数.

任意 (可测) 集合 $A\in\mathscr{F}$ 的示性函数 $I_A(\omega)$, 是随机变量最简单的例子.

表示为

$$\xi(\omega)=\sum_{i=1}^{\infty}x_i I_{A_i}(\omega) \tag{2}$$

的随机变量, 其中 $\sum A_i=\Omega, A_i\in\mathscr{F}$, 称做离散型随机变量. 假如 (2) 式中的和含有限项, 则相应的随机变量称做简单的.

按照第一章 §4 的解释, 可以说随机变量试验的某一数量特征, 其取值依赖于 "偶然" ω. 这时, 可测性 (1) 的要求之所以重要, 其原因是: 如果在 $(\Omega,\mathscr{F})$ 上给定了概率, 则事件 $\{\omega:\xi(\omega)\in B\}$ 的概率才有意义, 其中 $\{\omega:\xi(\omega)\in B\}$ 表示 "随机变量 ξ 的值属于某一给定的博雷尔集合 B".

于是, 我们给出如下定义.

定义 2 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的概率测度 P_ξ 连同

$$P_\xi(B)=\mathbf{P}\{\omega:\xi(\omega)\in B\},B\in\mathscr{B}(\mathbb{R}),$$

称做随机变量 ξ 在 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的概率分布.

定义 3 函数

$$F_\xi(x)=\mathbf{P}\{\omega:\xi(\omega)\leqslant x\},x\in\mathbb{R},$$

称做随机变量 ξ 的分布函数.

对于离散型随机变量 ξ, 测度 P_ξ 集中在有限或可数个点上, 并且可以表示为

$$P_\xi(B)=\sum_{\{k:x_k\in B\}}p(x_k), \tag{3}$$

其中 $p(x_k)=\mathbf{P}\{\xi=x_k\}=\Delta F_\xi(x_k)$.

显然, 也有相反的结论: 如果 P_ξ 可以表示为 (3) 式, 则 ξ 是离散型随机变量.

随机变量 ξ 称做连续的, 如果其分布函数 $F_\xi(x)$ 对 $x\in\mathbb{R}$ 连续.

随机变量 ξ 称做绝对连续的[①], 如果存在非负函数 $f=f_\xi(x)$, 使

$$F_\xi(x)=\int_{-\infty}^{x}f_\xi(y)\mathrm{d}y,x\in\mathbb{R} \tag{4}$$

(积分可以理解为黎曼积分, 不过在更一般的情形下, 应理解为勒贝格积分; 参见下面 §6), 其中函数 $f=f_\xi(x)$ 称做密度.

2. 函数 $\xi=\xi(\omega)$ 为 $\mathscr{F}$- 可测的充分和必要条件 判断函数 $\xi=\xi(\omega)$ 是否随机变量, 需要对于所有集合 $B\in\mathscr{F}$ 验证性质 (1) 是否成立. 下面的引理表明, 对于这样集合类的 "测试" 可以简化.

引理 1 设 $\mathscr{E}$ 是某一集系, 且 $\sigma(\mathscr{E})=\mathscr{B}(\mathbb{R})$. 函数 $\xi=\xi(\omega)$ 为 $\mathscr{F}$- 可测的充分和必要条件是, 对于一切 $E\in\mathscr{E}$, 有

$$\{\omega:\xi(\omega)\in E\}\in\mathscr{F}. \tag{5}$$

证明 必要性显然. 为证明充分性仍利用适当集合原理 (§2).

设 $\mathscr{D}$ 是满足 $\xi^{-1}(D)\in\mathscr{F},D\in\mathscr{B}(\mathbb{R})$ 的博雷尔集系. 不难验证 "取逆像" 运算, 保持并、交与补的集合运算不变:

$$\begin{gathered}\xi^{-1}\left(\bigcup_\alpha B_\alpha\right)=\bigcup_\alpha\xi^{-1}(B_\alpha),\\ \xi^{-1}\left(\bigcap_\alpha B_\alpha\right)=\bigcap_\alpha\xi^{-1}(B_\alpha),\\ \overline{\xi^{-1}(B_\alpha)}=\xi^{-1}(\overline{B}_\alpha).\end{gathered} \tag{6}$$

① 在我国文献中, 亦常称绝对随机变量为连续型随机变量.—— 译者

由此可见, 集系 $\mathscr{D}$ 是 $\sigma-$ 代数. 因此

$$\mathscr{E} \subseteq \mathscr{D} \subseteq \mathscr{B}(\mathbb{R}),$$

故

$$\sigma(\mathscr{E}) \subseteq \sigma(\mathscr{D}) = \mathscr{D} \subseteq \mathscr{B}(\mathbb{R}).$$

由于 $\sigma(\mathscr{E}) = \mathscr{B}(\mathbb{R})$, 从而 $\mathscr{D} = \mathscr{B}(\mathbb{R})$. □

系　*函数 $\xi = \xi(\omega)$ 为随机变量的充分和必要条件是, 对于任何 $x \in \mathbb{R}$, 有*

$$\{\omega : \xi(\omega) < x\} \in \mathscr{F}$$

或

$$\{\omega : \xi(\omega) \leqslant x\} \in \mathscr{F}.$$

由如下事实可以立即得到系的证明: 每一个集系

$$\mathscr{E}_1 = \{x : x < c, c \in \mathbb{R}\},$$
$$\mathscr{E}_2 = \{x : x \leqslant c, c \in \mathbb{R}\}$$

都产生 $\sigma-$ 代数 $\mathscr{B}(\mathbb{R})$, 即 $\sigma(\mathscr{E}_1) = \sigma(\mathscr{E}_2) = \mathscr{B}(\mathbb{R})$ (参见 §2),

基于下面的引理, 可以用其他随机变量的函数来构造新的随机变量.

引理 2　*设 $\varphi = \varphi(x)$ 是博雷尔函数, 而 $\xi = \xi(\omega)$ 是一随机变量. 那么, 复合函数 $\eta = \varphi \circ \xi$, 即函数 $\eta(\omega) = \varphi(\xi(\omega))$ 也是随机变量.*

证明　引理的结论由如下事实可得: 对于 $B \in \mathscr{B}(\mathbb{R})$,

$$\{\omega : \eta(\omega) \in B\} = \{\omega : \varphi(\xi(\omega)) \in B\} = \{\omega : \xi(\omega) \in \varphi^{-1}(B)\} \in \mathscr{F}, \tag{7}$$

因为 $\varphi^{-1}(B) \in \mathscr{B}(\mathbb{R})$. □

这样, 由于 $x^n, x^+, x^-, |x|$ 是博雷尔函数, 可见如果 ξ 是随机变量, 则 $\xi^n, \xi^+ = \max\{\xi, 0\}, \xi^- = -\min\{\xi, 0\}, |\xi|$ 也都是随机变量 (练习题 3).

3. 广义随机变量[①]　从给定的随机变量组 $\{\xi_n\}$ 出发, 可以建立随机变量新的函数, 例如,

$$\sum_{n=1}^{\infty} |\xi_n|, \overline{\lim_n} \xi_n, \underline{\lim_n} \xi_n, \cdots$$

注意, 这些函数一般已经在扩充数轴 $[-\infty, \infty]$ 上取值. 因此, 最好将可测函数类加以扩充, 使之也可以取 $\pm\infty$ 为值.

[①] Расширенная случайная величина　(extendid random vriable), 亦可译为 "扩充随机变量" —— 译者

定义 4 称定义在 $(\Omega,\mathscr{F})$ 上、取值于 $\overline{\mathbb{R}}=[-\infty,\infty]$ 上的函数 $\xi=\xi(\omega)$ 为广义随机变量, 如果对于任何博雷尔集 $B\in\mathscr{B}(\overline{\mathbb{R}})$ (σ- 代数 $\mathscr{B}(\overline{\mathbb{R}})=\sigma(\mathscr{B}(\mathbb{R}),\pm\infty)$) 满足条件 (1).

下面的定理虽然简单, 但是在建立勒贝格积分时是关键 (§6).

定理 1 a) 对于任意 (包括广义) 随机变量 $\xi=\xi(\omega)$, 存在简单随机变量列 $\xi_1,\xi_2,\cdots$, 使 $|\xi_n|\leqslant|\xi|$, 且对于一切 $\omega\in\Omega$, 当 $n\to\infty$ 时 $\xi_n(\omega)\to\xi(\omega)$.

b) 在上述条件下, 如果 $\xi(\omega)\geqslant 0,\omega\in\Omega$, 则存在简单随机变量序列 $\xi_1,\xi_2,\cdots$, 使对于一切 $\omega\in\Omega$, 当 $n\to\infty$ 时 $\xi_n(\omega)\uparrow\xi(\omega)$.

证明 首先证明第 2 个命题. 对于 $n=1,2,\cdots$, 设

$$\xi_n(\omega)=\sum_{k=1}^{n2^n}\frac{k-1}{2^n}I_{\left\{\frac{k-1}{2^n}\leqslant\xi(\omega)<\frac{k}{2^n}\right\}}(\omega)+I_{\{\xi(\omega)\geqslant n\}}(\omega).$$

可以直接验证, 对于一切 $\omega\in\Omega$, 所建立的序列 $\xi_n(\omega)\uparrow\xi(\omega)$. 如果注意到, ξ 可以表示为 $\xi=\xi^+-\xi^-$, 则由已经证明的命题 b), 可以立即得到命题 a). □

现在证明, 广义随机变量类关于逐项收敛封闭. 为此, 首先指出, 如果 $\xi_1,\xi_2,\cdots$ 是广义随机变量序列, 则 $\sup\ \xi_n,\inf\ \xi_n,\overline{\lim}\xi_n$ 和 $\underline{\lim}\xi_n$ 也是随机变量 (有可能是广义的). 由

$$\{\omega:\sup\xi_n>x\}=\bigcup_n\{\omega:\xi_n>x\}\in\mathscr{F},$$

$$\{\omega:\inf\ \xi_n<x\}=\bigcup_n\{\omega:\xi_n<x\}\in\mathscr{F};$$

$$\overline{\lim}\xi_n=\inf_n\ \sup_{m\geqslant n}\xi_m,\underline{\lim}\xi_n=\sup_n\ \inf_{m\geqslant n}\xi_m,$$

可以直接得到上面所指出的性质.

定理 2 如果 $\xi_1,\xi_2,\cdots$ 是扩充随机变量序列, 而 $\xi(\omega)=\lim\xi_n(\omega),\omega\in\Omega$, 则 $\xi=\xi(\omega)$ 也是广义随机变量.

证明 定理的证明, 可以立即由上面的结果和下面的关系式得到:

$$\begin{aligned}\{\omega:\xi_n<x\}&=\{\omega:\lim\xi_n(\omega)<x\}\\&=\{\omega:\overline{\lim}\xi_n(\omega)=\underline{\lim}\xi_n(\omega)\}\cap\{\overline{\lim}\xi_n(\omega)<x\}\\&=\Omega\cap\{\overline{\lim}\xi_n(\omega)<x\}=\{\overline{\lim}\xi_n(\omega)<x\}\in\mathscr{F}.\end{aligned}$$

4. 随机变量序列之和、差、积、商的极限 我们现在讨论随机变量的简单函数的一些性质, 这些变量定义在可测空间 $(\Omega,\mathscr{F})$ 上、且有可能取值于扩充数轴 $\overline{\mathbb{R}}=[-\infty,\infty]$ *).

*) 以下关于 $\overline{\mathbb{R}}$ 中的算术运算, 作如下通常的约定: 若 $a\in\mathbb{R}$, 则 $a\pm\infty=\pm\infty, a/\pm\infty=0$; 若 $a<0$, 则 $a\times\infty=-\infty$; 此外, $0\times(\pm\infty)=0,\infty+\infty=\infty,-\infty-\infty=-\infty$.

如果 ξ 和 η 是两个随机变量, 则 $\xi+\eta, \xi-\eta, \xi\eta$ 和 ξ/η 也是随机变量 (假设有关运算有意义, 即不出现形如 $\infty-\infty, \infty/\infty, 0/0$ 等不定式).

设 $\{\xi_n\}$ 和 $\{\eta_n\}$ 是两个随机变量序列, 分别收敛于 ξ 和 η (见定理 1). 那么

$$\xi_n \pm \eta_n \to \xi \pm \eta,$$

$$\xi_n \eta_n \to \xi\eta,$$

$$\frac{\xi_n}{\eta_n + \dfrac{1}{n} I_{\{\eta_n=0\}}(\omega)} \to \frac{\xi}{\eta}.$$

这些关系式左侧的各个函数都是简单随机变量. 因此, 由定理 2 知 $\xi \pm \eta, \xi\eta, \xi/\eta$ 都是随机变量.

5. 随机变量的函数 设 $\xi=\xi(\omega)$ 是随机变量. 考虑 $\mathscr{F}$ 中形如 $\{\omega : \xi(\omega) \in B\}, B \in \mathscr{B}(\mathbb{R})$ 的集合, 它们构成 $\sigma-$ 代数, 称为由随机变量 ξ 生成的 $\sigma-$ 代数, 记作 $\mathscr{F}_\xi$ 或 $\sigma(\xi)$.

如果 φ 是某一博雷尔函数, 则由引理 2 知函数 $\eta=\varphi\circ\xi$ 也是随机变量, 并且 $\mathscr{F}_\xi-$ 可测: $\{\omega : \eta(\omega) \in B\} \in \mathscr{F}_\xi, B \in \mathscr{B}(\mathbb{R})$ (见 (7) 式). 实际上, 相反的结果也成立.

定理 3 设 $\eta=\eta(\omega)$ 是 $\mathscr{F}_\xi-$ 可测随机变量. 那么, 存在某一博雷尔函数 φ, 使对于每一个 $\omega \in \Omega$, 有 $\eta=\varphi\circ\xi$, 即 $\eta(\omega)=\varphi(\xi(\omega))$.

证明 设 Φ_ξ 是一切 $\mathscr{F}_\xi-$ 可测函数 $\eta=\eta(\omega)$ 类, 而 $\overline{\Phi}_\xi$ 是可以表示为 $\varphi\circ\eta$ 的 $\mathscr{F}_\xi-$ 可测函数类, 其中 φ 是某一博雷尔函数. 显然, $\overline{\Phi}_\xi \subseteq \Phi_\xi$. 定理的结论表明, 实际上是 $\overline{\Phi}_\xi = \Phi_\xi$.

设 $A \in \mathscr{F}_\xi$ 和 $\eta(\omega)=I_A(\omega)$. 现在证明 $\eta \in \overline{\Phi}_\xi$. 事实上, 若 $A \in \mathscr{F}_\xi$, 则存在 $B \in \mathscr{B}(\mathbb{R})$, 使 $A=\{\omega : \xi(\omega) \in B\}$. 记

$$\chi_B(x) = \begin{cases} 1, & 若\ x \in B, \\ 0, & 若\ x \notin B. \end{cases}$$

那么 $I_A(\omega)=\chi_B(\xi(\omega)) \in \overline{\Phi}_\xi$. 由此可见, 任何 $\mathscr{F}_\xi-$ 可测简单函数

$$\sum_{i=1}^{n} c_i I_{A_i}(\omega), A_i \in \mathscr{F}_\xi,$$

仍然属于 $\overline{\Phi}_\xi$.

现在设 η 是任意 $F_\xi-$ 可测简单函数. 根据定理 1, 存在 $\mathscr{F}_\xi-$ 可测的简单函数序列 $\{\eta_n\}$, $\eta_n=\eta_n(\omega)$, 使得当 $n\to\infty$ 时 $\eta_n(\omega)\to\eta(\omega), \omega \in \Omega$. 由上面证明的结果知, 存在博雷尔函数 $\varphi_n=\varphi_n(x)$, 使 $\eta_n=\varphi_n(\xi(\omega))$. 在这种情况下当 $n\to\infty$ 时 $\varphi_n(\xi(\omega))\to\eta(\omega), \omega\in\Omega$.

记 $B=\{x\in\mathbb{R}:\lim_n\varphi_n(x)$存在$\}$. 此集合是博雷尔集. 所以函数

$$\varphi(x)=\begin{cases}\lim\limits_n\varphi_n(x), & 若\ x\in B,\\ 0, & 若\ x\notin B\end{cases}$$

仍然是博雷尔函数 (见练习题 6).

那么, 显然对于所有 $\omega\in\Omega$, 有

$$\eta(\omega)=\lim_n\varphi_n(\xi(\omega))=\varphi(\xi(\omega)).$$

于是, $\overline{\Phi}_\xi=\Phi_\xi$. □

6. 阶梯随机变量 考虑可测空间 $(\Omega,\mathscr{F})$, 和空间 Ω 的有限或可数分割: $\mathscr{D}=\{D_1,D_2,\cdots\},D_i\in\mathscr{F},\sum_i D_i=\Omega$. 组成含空集 $\varnothing$ 和 $\sum_\alpha D_\alpha$ 的代数 $\mathscr{A}$, 其中和式含有限或可数项. 显然, 集系 $\mathscr{A}$ 是单调类, 因此根据 §2 引理 2, 代数 $\mathscr{A}$ 同时也是 $\sigma-$ 代数, 记作 $\sigma(\mathscr{D})$, 称为由 $\mathscr{D}$ 生成的 $\sigma-$ 代数. 显然 $\sigma(\mathscr{D})\subseteq F$.

引理 3 *设 $\xi=\xi(\omega)$ 是 $\sigma(\mathscr{D})-$ 可测随机变量, 则 ξ 可以表示为:*

$$\xi(\omega)=\sum_{k=1}^{\infty}x_kI_{D_k}(\omega), \tag{8}$$

其中 $x_k\in\mathbb{R},k\geqslant 1$, 即 $\xi(\omega)$ 在分割的元素 $D_k,k\leqslant 1$ 上为常数.

证明 任取分割的一个集合 D_k, 证明 $\sigma(\mathscr{D})-$ 可测函数 ξ 在 D_k 上为常数.

为此, 设

$$x_k=\sup[c:D_k\cap\{\omega:\xi(\omega)<c\}=\varnothing].$$

由于

$$\{\omega:\xi(\omega)<x_k\}=\bigcup_{r<x_k}\{\omega:\xi(\omega)<r\},$$

而 $r<x_k$ 是有理数, 则 $D_k\bigcap\{\omega:\xi(\omega)<x_k\}=\varnothing$.

现在设 $c>x_k$, 则 $D_k\bigcap\{\omega:\xi(\omega)<c\}\neq\varnothing$. 由于集合 $\{\omega:\xi(\omega)<c\}$ 具有 $\sum_\alpha D_\alpha$ 的形式, 其中对于有限或可数个下标求和, 故

$$D_k\cap\{\omega:\xi(\omega)<c\}=D_k.$$

由此可见, 对于一切 $c>x_k$,

$$D_k\cap\{\omega:\xi(\omega)\geqslant c\}=\varnothing;$$

由于

$$\{\omega:\xi(\omega)>x_k\}=\bigcup_{r>x_k}\{\omega:\xi(\omega)\geqslant r\},$$

其中对于有限或可数个下标求和, 故

$$D_k \cap \{\omega : \xi(\omega) > x_k\} = \varnothing.$$

于是, $D_k \bigcap \{\omega : \xi(\omega) \neq x_k\} = \varnothing$. 从而

$$D_k \subseteq \{\omega : \xi(\omega) = x_k\},$$

而这正是需要证明的. □

7. 练习题

1. 证明随机变量 ξ 连续的充分和必要条件是: 对于一切 $x \in \mathbb{R}, \mathbf{P}\{\xi = x\} = 0$.
2. 设 $|\xi|$ 为 $\mathscr{F}-$ 可测, 问 ξ 是否也 $\mathscr{F}-$ 可测?
3. 证明函数 $x^n, x^+ = \max(x, 0), x^- = -\min(x, 0), |x| = x^+ + x^-$ 是博雷尔函数.
4. 如果 ξ 和 η 都 $\mathscr{F}-$ 可测, 则 $\{\omega : \xi(\omega) = \eta(\omega)\} \in \mathscr{F}$.
5. 设 ξ 和 η 是 $(\Omega, \mathscr{F})$ 上的两个随机变量, 而集合 $A \in \mathscr{F}$. 证明, 函数

$$\zeta(\omega) = \xi(\omega) I_A + \eta(\omega) I_{\overline{A}}$$

也是随机变量.

6. 设 $\xi_1, \cdots, \xi_n$ 是随机变量, 而 $\varphi(x_1, \cdots, x_n)$ 是博雷尔函数. 证明 $\varphi(\xi_1(\omega), \cdots, \xi_n(\omega))$ 也是随机变量.

7. 设 ξ 和 η 是以 $1, 2, \cdots, N$ 为值的两个随机变量, 且 $\mathscr{F}_\xi = \mathscr{F}_\eta$. 证明存在数列的 $(1, 2, \cdots, N)$ 的排列 $(i_1, i_2, \cdots, i_N)$, 使对于每一个 $j = 1, 2, \cdots, N$, 集合 $\{\omega : \xi = j\}$ 与 $\{\omega : \eta = i + j\}$ 相等.

8. 举一随机变量 ξ 的例, 使其分布函数有密度 $f(x)$, 但是当 $x \to \infty$ 时 $f(x)$ 的极限 $\lim_{x \to \infty} f(x)$ 不存在, 从而 $f(x)$ 在无穷大不为 0.

9. 设 ξ 和 η 是有界随机变量 $(|\xi| \leqslant c_1, |\eta| \leqslant c_2)$. 证明, 若对于一切 $m, n \geqslant 1$, 有

$$\mathbf{E}\xi^m \eta^n = \mathbf{E}\xi^m \times \mathbf{E}\eta^n,$$

则 ξ 和 η 独立.

10. 设 ξ 和 η 是随机变量, 其分布函数 $F_\xi = F_\eta$. 证明, 若对 $x \in \mathbb{R}, \{\omega : \xi(\omega) = x\} \neq \varnothing$, 则存在 $y \in \mathbb{R}$, 使 $\{\omega : \xi(\omega) = x\} = \{\omega : \eta(\omega) = y\}$.

11. 设 E 是 $\mathbb{R}$ 上的有限或可数集合,ξ 是映射: $\Omega \to E$. 证明 ξ 是 $(\Omega, \mathscr{F})$ 上是随机变量, 当且仅当对于每一个 $x \in E, \{\omega : \xi(\omega) = x\} \in \mathscr{F}$.

§5. 随机元

1. 随机函数、随机向量和随机过程 除随机变量外, 在概率论及其应用中, 还研究更一般性质的随机对象. 例如, 随机点、随机向量、随机函数、随机过程、随机

场、随机集合、随机测度等等. 因此, 希望有一个关于任意性质的随机对象的一般概念.

定义 1 设 $(\Omega, \mathscr{F})$ 和 $(E, \mathscr{E})$ 是两个可测空间. 定义在 Ω 上取值于 E 的函数 $X = X(\omega)$, 称做 (取值于 E 的) $\mathscr{F}/\mathscr{E}-$可测函数或随机元, 如果对于任何 $B \in \mathscr{E}$, 有

$$\{\omega : X(\omega) \in B\} \in \mathscr{F}. \tag{1}$$

取值于 E 的随机元, 亦常称为 $E-$值随机变量.

下面讨论这一定义的一些特殊情形.

假如 $(E, \mathscr{E}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, 则随机元的定义与随机变量的定义一致 (§4).

假如 $(E, \mathscr{E}) = (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, 则随机元 $X(\omega)$ 是 $\mathbb{R}^n$ 中的 "随机点". 如果 π_k 是 $\mathbb{R}^n$ 在第 k 坐标轴上的射影, 则 $X(\omega)$ 可以表示为

$$X(\omega) = (\xi_1(\omega), \cdots, \xi_n(\omega)), \tag{2}$$

其中 $\xi_k = \pi_k \circ X$.

由条件 (1) 可见, ξ_k 是普通的随机变量. 事实上, 对于任意 $B \in \mathscr{B}(\mathbb{R})$, 由于 $\mathbb{R} \times \cdots \times \mathbb{R} \times B \times \mathbb{R} \times \cdots \times \mathbb{R} \in \mathscr{B}(\mathbb{R}^n)$, 有

$$\begin{aligned} &\{\omega : \xi_k(\omega) \in B\} \\ &= \{\omega : \xi_1(\omega) \in \mathbb{R}, \cdots, \xi_{k-1}(\omega) \in \mathbb{R}, \xi_k(\omega) \in B, \xi_{k+1}(\omega) \in \mathbb{R}, \cdots, \xi_n(\omega) \in \mathbb{R}\} \\ &= \{\omega : X(\omega) \in (\mathbb{R} \times \cdots \times \mathbb{R} \times B \times \mathbb{R} \times \cdots \times \mathbb{R})\} \in \mathscr{F}. \end{aligned}$$

定义 2 我们把任意有序随机变量组 $(\eta_1(\omega), \cdots, \eta_n(\omega))$, 称做 $n-$维随机向量.

按照这一定义, 任意在 $\mathbb{R}^n$ 取值的随机元 $X(\omega)$ 是 $n-$ 维随机向量. 反过来也对, 任何 $n-$ 维随机向量 $X(\omega) = (\xi_1(\omega), \cdots, \xi_n(\omega))$, 是 $\mathbb{R}^n$ 中的随机元. 事实上, 如果 $B_k \in \mathscr{B}(\mathbb{R}), k = 1, \cdots, n$, 则

$$\{\omega : X(\omega) \in (B_1 \times \cdots \times B_n)\} = \prod_{k=1}^{n} \{\omega : \xi_k(\omega) \in B_k\} \in \mathscr{F}.$$

含 $B_1 \times \cdots \times B_n$ 的最小 $\sigma-$ 代数等于 $\mathscr{B}(\mathbb{R}^n)$. 那么, 由 §4 中引理 1 的明显推广立即可得, 对于任意 $B \in \mathscr{B}(\mathbb{R}^n)$, 集合 $\{\omega : X(\omega) \in B\} \in \mathscr{F}$.

设 $(E, \mathscr{E}) = (\mathbf{Z}, \mathscr{B}(\mathbf{Z}))$, 其中 $\mathbf{Z}$ 是复数 $z = x + iy, x, y \in \mathbb{R}$ 的集合, 而 $\mathscr{B}(\mathbf{Z})$ 含形如 $\{z : z = x + iy, a_1 < x \leqslant b_1, a_2 < y \leqslant b_2\}$ 集合的最小 $\sigma-$ 代数. 由以上的讨论, 可见复值随机变量 $Z(\omega)$ 可以表示为 $Z(\omega) = X(\omega) + iY(\omega)$, 其中 $X(\omega)$ 和 $Y(\omega)$ 是随机变量. 因此, $Z(\omega)$ 亦称为复随机变量.

设 $(E, \mathscr{E}) = (\mathbb{R}^T, \mathscr{B}(\mathbb{R}^T))$, 其中 T 是数轴的某一子集. 这时, 任何可以表示为 $X = (\xi_t)_{t \in T}, \xi_t = \pi_t \circ X$ 的随机元 $X = X(\omega)$, 称做定义在时间区间 T 上的随机函数.

像随机向量一样, 任何随机函数同时也是随机过程. 见下面的定义.

定义 3　设 T 是数轴的某一子集. 随机变量组 $X=(\xi_t)_{t\in T}$ 称做时间区间T上的**随机过程**.

对于 $T=\{1,2,\cdots\}$ 的情形, $X=(\xi_1,\xi_2,\cdots)$ 称为**离散时间随机过程**或**随机序列**.

对于 $T=[0,1],(-\infty,\infty),[0,\infty),\cdots$ 的情形, $X=(\xi_t)_{t\in T}$ 称为**连续时间随机过程**.

利用 $\sigma-$ 代数 $\mathscr{B}(\mathbb{R}^T)$ 的构造 (§2), 不难证明, (在定义 3 的意义下的) 任何随机过程 $X=(\xi_t)_{t\in T}$, 同时也是随机函数 (值域为 $\mathbb{R}^T$ 的随机元).

定义 4　设 $X=(\xi_t)_{t\in T}$ 是随机过程. 对于每个固定的 $\omega\in\Omega$, 函数 $(\xi_t(\omega))_{t\in T}$ 称做过程对应于结局 ω 的**实现**或**轨道**.

与 §4 的定义 2 类似, 自然地引出下面的定义.

定义 5　设 $X=(\xi_t)_{t\in T}$ 是随机过程.

1) $(\mathbb{R}^T,\mathscr{B}(\mathbb{R}^T))$ 上的概率测度 P_X, 其中

$$P_X(B)=\mathbf{P}\{\omega:X(\omega)\in B\},\ B\in\mathscr{B}(\mathbb{R}^T),$$

称做过程 X 的**概率分布**.

2) 对于 $t_1<t_2\cdots<t_n,t_i\in T$, 概率

$$P_{t_1,\cdots,t_n}(B)=\mathbf{P}\{\omega:(\xi_{t_1},\cdots,\xi_{t_n})\in B\},B\in\mathscr{B}(\mathbb{R}^n),$$

称做过程 X **有限维概率** (或**有限维概率分布**).

3) 对于 $t_1<t_2\cdots<t_n,t_i\in T$ 函数

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)\equiv\mathbf{P}\{\omega:\xi_{t_1}\leqslant x_1,\cdots,\xi_{t_n}\leqslant x_n\}$$

称做过程 $X=(\xi_t)_{t\in T}$ 的**有限维概率分布函数**.

设 $(E,\mathscr{E})=(C,\mathscr{B}_0(C))$, 其中 C 是区间 $T=[0,1]$ 上连续函数 $x=(x_t)_{t\in T}$ 的空间, 其中 $\sigma-$ 代数 $\mathscr{B}_0(C)$ 是由开集生成的 (§2). 下面证明, 空间 $(C,\mathscr{B}_0(C))$ 的任意随机元 X, 同时也是在定义 3 意义上的随机过程 (且具有连续轨道).

实际上根据 §2, 集合 $A=\{x\in C:x_t<a\}$ 是 $\mathscr{B}_0(C)$ 中的开集. 因此

$$\{\omega:\xi_t(\omega)<a\}=\{\omega:X(\omega)\in A\}\in\mathscr{F}.$$

另一方面, 设 $X=(\xi_t(\omega))_{t\in T}$ 是 (在定义 3 意义上的) 随机过程, 对于每个 $\omega\in\Omega$, 其轨道是连续函数. 根据 §2 的 (17) 式

$$\{x\in C:x\in S_\rho(X^0(\omega))\}=\bigcap_{t_k}\{x\in C:|x_{t_k}-x^0_{t_k}|<\rho\},$$

其中 t_k 是线段 [0,1] 上的有理点. 因此,

$$\{\omega : X(\omega) \in S_\rho(X^0(\omega))\} = \bigcap_{t_k} \{\omega : |\xi_{t_k}(\omega) - \xi^0_{t_k}(\omega)| < \rho\} \in \mathscr{F}.$$

于是, 对于任何 $\{\omega : X(\omega) \in B\} \in \mathscr{F}$, 有 $B \in \mathscr{B}_0(C)$.

通过类似的讨论也可以证明, 在 §2 (第 7 小节) 中引进的空间 $(D, \mathscr{B}_0(D))$ 上的随机元, 可以视为随机过程, 其轨道属于无第二类间断点的函数空间. 并且逆命题成立.

2. 随机元的独立性 设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, 而 $(E_\alpha, \mathscr{E}_\alpha)$ 是可测空间, 其中 α 的值域是某 (任意) 集合 $\mathfrak{U}$.

定义 6 称 $\mathscr{F}/\mathscr{E}_\alpha-$ 可测函数 $(X_\alpha(\omega)), \alpha \in \mathfrak{U}$, 独立 (或全体独立), 如果对于任何有限下标数组 $\alpha_1, \cdots, \alpha_n$, 随机元 $X_{\alpha_1}, \cdots, X_{\alpha_n}$ 独立, 即

$$\mathbf{P}\{X_{\alpha_1} \in B_{\alpha_1}, \cdots, X_{\alpha_n} \in B_{\alpha_n}\} = \mathbf{P}\{X_{\alpha_1} \in B_{\alpha_1}\} \cdots \mathbf{P}\{X_{\alpha_n} \in B_{\alpha_n}\}, \tag{3}$$

其中 $B_\alpha \in \mathscr{E}_\alpha$.

设 $\mathfrak{U} = \{1, 2, \cdots, n\}, \xi_\alpha$ 是随机变量, $\alpha \in \mathfrak{U}$, 而

$$F_\xi(x_1, \cdots, x_n) = \mathbf{P}\{\xi_1 \leqslant x_1, \cdots, \xi_n \leqslant x_n\}$$

是向量 $(\xi_1, \cdots, \xi_n)$ 的 $n-$ 维分布函数. 设 $F_{\xi_i}(x_i)$ 是随机变量 $\xi_i, i = 1, \cdots, n$, 的分布函数.

定理 随机变量 $\xi_1, \cdots, \xi_n$ 独立的充分和必要条件是, 对于所有向量 $(x_1, \cdots, x_n) \in \mathbb{R}^n$, 有

$$F_\xi(x_1, \cdots, x_n) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n). \tag{4}$$

证明 必要性显然. 为证明充分性, 设 $a = (a_1, \cdots, a_n), b = (b_1, \cdots, b_n)$ 是任意向量, 则

$$\mathbf{P}_\xi(a, b] = \mathbf{P}\{a_1 < \xi_1 \leqslant b_1, \cdots, a_n < \xi_n \leqslant b_n\},$$
$$\mathbf{P}_{\xi_i}(a_i, b_i] = \mathbf{P}\{a_i < \xi_i \leqslant b_i\}.$$

那么, 由 §3 的 (7) 式, 以及 (4) 式, 得

$$\mathbf{P}_\xi(a, b] = \prod_{i=1}^n [F_{\xi_i}(b_i) - F_{\xi_i}(a_i)] = \prod_{i=1}^n P_{\xi_i}(a_i, b_i].$$

从而,

$$\mathbf{P}\{\xi_1 \in I_1, \cdots, \xi_n \in I_n\} = \prod_{i=1}^n \mathbf{P}\{\xi_i \in I_i\}, \tag{5}$$

其中 $I_i = (a_i, b_i]$.

固定 $I_2, \cdots, I_n$, 对于任意 $B_1 \in \mathscr{B}(\mathbb{R})$, 有

$$\mathbf{P}\{\xi_1 \in B_1, \xi_2 \in I_2, \cdots, \xi_n \in I_n\} = \mathbf{P}\{\xi_1 \in B_1\} \prod_{i=2}^{n} \mathbf{P}\{\xi_i \in I_i\}, \tag{6}$$

设 $\mathscr{M}$ 是 $\mathscr{B}(\mathbb{R})$ 中满足 (6) 式的集合的全体 (§2, "适当集合原理"). 由不相交的形如 $I_1 = (a_1, b_1]$ 的区间形成的代数 $\mathscr{A}$ 显然属于 $\mathscr{M}$. 因此 $\mathscr{A} \subseteq \mathscr{M} \subseteq \mathscr{B}(\mathbb{R})$. 由概率测度的可数可加性 (从而也由其连续性), 可见集系 $\mathscr{M}$ 是单调类. 因此 (见 §2, 第 1 小节)

$$\mu(\mathscr{M}) \subseteq \mathscr{M} \subseteq B(\mathbb{R}).$$

然而. 根据 §2 的定理 1, $\mu(\mathscr{A}) = \sigma(\mathscr{A}) = \mathscr{B}(\mathbb{R})$, 故 $\mathscr{M} = \mathscr{B}(\mathbb{R})$.

于是 (6) 式得证. 现在固定 $B_1, I_3, \cdots, I_n$, 并将 I_2 换成 B_2, 运用与证明 (6) 式同样的方法继续这一过程, 则显然可以得到需要的等式:

$$\mathbf{P}\{\xi_1 \in B_1, \cdots, \xi_n \in B_n\} = \mathbf{P}\{\xi_1 \in B_1\} \cdots \mathbf{P}\{\xi_n \in B_n\},$$

其中 $B_i \in \mathscr{B}(\mathbb{R})$. □

3. 练习题

1. 设 $\xi_1, \cdots, \xi_n$ 是离散型随机变量, 证明它们独立的充分和必要条件是, 对于任意实数 $x_1, \cdots, x_n$, 有

$$\mathbf{P}\{\xi_1 = x_1, \cdots, \xi_n = x_n\} = \prod_{i=1}^{n} \mathbf{P}\{\xi_i = x_i\}.$$

2. 证明任何随机函数 $X(\omega) = (\xi_t(\omega))_{t \in T}$ (在定义 3 的意义上) 是随机过程, 而且反之亦然.

3. 设 $X_1, \cdots, X_n$ 是分别在 $(E_1, \mathscr{E}_1), \cdots, (E_n, \mathscr{E}_n)$ 取值的随机元. 其次, 设 $(E_1', \mathscr{E}_1'), \cdots, (E_n', \mathscr{E}_n')$ 是可测空间, 而 $g_1, \cdots, g_n$ 相应为 $\mathscr{E}_1/\mathscr{E}_1', \cdots, \mathscr{E}_n/\mathscr{E}_n'-$ 可测函数. 证明, 若 $X_1, \cdots, X_n$ 独立, 则 $g_1 \circ X_1, \cdots, g_n \circ X_n$ 也独立.

4. 设 $X_1, X_n, \cdots$ 是可交换无限随机变量序列 (即这样的序列, 每一组随机变量, 由具有不同下标的 k 个元素构成, 例如, $X_{i_1}, \cdots, X_{i_k}$ 只与 k 有关, 与 $i_1, \cdots, i_k$ 具体值无关, 其中 $i_1, \cdots, i_k$ 两两不等; 对照 §1 的练习题 11). 证明, 若 $\mathbf{E}X_n^2 < \infty, n \geqslant 1$, 则协方差 $\operatorname{cov}(X_1, X_2) \geqslant 0$.

5. 设 ξ, η, ζ 是独立随机变量, 证明随机变量 $\xi + \eta$ 与 ζ^2 独立.

6. 设由随机变量 $\xi_1, \cdots, \xi_m, \eta_1, \cdots, \eta_n$ 组成随机向量 $X = (\xi_1, \cdots, \xi_m)$ 和 $Y = (\eta_1, \cdots, \eta_n)$. 假设满足下列条件:

(i) 随机变量 $\xi_1, \cdots, \xi_m$ 独立;

(ii) 随机变量 $\eta_1, \cdots, \eta_n$ 独立;

(iii) 随机向量 X 和 Y, 作为分别取值于 $\mathbb{R}^m$ 和 $\mathbb{R}^n$ 的随机元独立.

证明随机变量 $\xi_1,\cdots,\xi_m,\eta_1,\cdots,\eta_n$ 独立.

7. 假设有两个随机向量 $X=(\xi_1,\cdots,\xi_m)$ 和 $Y=(\eta_1,\cdots,\eta_n)$. 已知随机变量 $\xi_1,\cdots,\xi_m,\eta_1,\cdots,\eta_n$ 独立.

(i) 证明, 随机向量 X 和 Y, 作为随机元独立 (对照练习题 6).

(ii) 设 $f:\mathbb{R}^m\to\mathbb{R}, g:\mathbb{R}^n\to\mathbb{R}$ 是博雷尔函数, 证明随机变量 $f(\xi_1,\cdots,\xi_m)$ 和 $g(\eta_1,\cdots,\eta_n)$ 独立.

§6. 勒贝格积分. 数学期望

1. 引言与记号 在第一章 §4, 对于 $(\Omega,\mathscr{F},\mathbf{P})$ 是有限概率空间, $\xi=\xi(\omega)$ 是简单随机变量

$$\xi(\omega)=\sum_{k=1}^{n}x_kI_{A_k}(\omega) \tag{1}$$

的情形, 曾经定义了数学期望 $\mathbf{E}\xi$ 的概念. 对于任意概率空间 $(\Omega,\mathscr{F},\mathbf{P})$, 使用与简单随机变量的数学期望同样的概念. 具体地说, 根据定义, 设

$$\mathbf{E}\xi=\sum_{k=1}^{n}x_k\mathbf{P}(A_k). \tag{2}$$

这一定义, 像有限概率空间一样, (在 $\mathbf{E}\xi$ 的值与形如 (1) 的表现无关的意义上) 是适定的. 类似地可以证明数学期望的简单性质 (见第一章 §4 第 5 小节).

这一节的目的是: 给出任意随机变量数学期望 $\mathbf{E}\xi$ 的定义, 研究其性质. 按分析的观点, 数学期望 $\mathbf{E}\xi$ 是 $\mathscr{F}-$ 可测函数 $\xi=\xi(\omega)$ 对测度 $\mathbf{P}$ 的勒贝格积分. 对于数学期望, 除 $\mathbf{E}\xi$ 之外还使用如下记号

$$\int_\Omega\xi(\omega)\mathbf{P}(\mathrm{d}\omega),\quad \int_\Omega\xi\mathrm{d}\mathbf{P}.$$

2. 数学期望的定义 设 $\xi=\xi(\omega)$ 是非负随机变量. 构造一非负简单随机变量序列 $\{\xi_n\}_{n\geqslant1}$, 使对于每个 $\omega\in\Omega$, 当 $n\to\infty$ 时 $\xi_n(\omega)\uparrow\xi(\omega)$ (见 §4, 定理 1).

由于 $\mathbf{E}\xi_n\leqslant\mathbf{E}\xi_{n+1}$ (对照第一章 §4 第 5 小节的性质 3)), 则存在 $\lim_n\mathbf{E}\xi_n$, 其中的极限有可能以 $+\infty$ 为值.

定义 1 称

$$\mathbf{E}\xi=\lim_n\mathbf{E}\xi_n \tag{3}$$

为非负随机变量 $\xi=\xi(\omega)$ 的勒贝格积分, 或 $\xi=\xi(\omega)$ 的数学期望.

为使该定义是适定的, 需要证明其中的极限与逼近序列 $\{\xi_n\}$ 的选择无关. 换句话说, 需要证明, 若 $\xi_n\uparrow\xi,\eta_m\uparrow\xi$, 其中 $\{\eta_m\}$ 也是非负简单函数序列, 则

$$\lim_n\mathbf{E}\xi_n=\lim_m\mathbf{E}\eta_m. \tag{4}$$

引理 1 设 η 和 $\xi_n(n \geqslant 1)$ 是非负简单随机变量, 且

$$\xi_n \uparrow \xi \geqslant \eta.$$

那么,

$$\lim_n \mathbf{E}\xi_n \geqslant \mathbf{E}\eta. \tag{5}$$

证明 设 $\varepsilon > 0$ 且

$$A_n = \{\omega : \xi_n \geqslant \eta - \varepsilon\}.$$

显然,$A_n \uparrow \Omega$, 而

$$\xi_n = \xi_n I_{A_n} + \xi_n I_{\overline{A}_n} \geqslant \xi_n I_{A_n} \geqslant (\eta - \varepsilon) I_{A_n}.$$

因此, 由简单随机变量的数学期望的性质, 可见

$$\begin{aligned}\mathbf{E}\xi_n &\geqslant \mathbf{E}(\eta - \varepsilon) I_{A_n} = \mathbf{E}\eta I_{A_n} - \varepsilon \mathbf{P}(A_n) \\ &= \mathbf{E}\eta - \mathbf{E}\eta I_{\overline{A}_n} - \varepsilon \mathbf{P}(A_n) \geqslant \mathbf{E}\eta - C\mathbf{P}(\overline{A}_n) - \varepsilon,\end{aligned}$$

其中 $C = \max_\omega \eta(\omega)$. 由于 $\varepsilon > 0$ 是任意的, 得所要证明的不等式 (5). □

由这一引理可见

$$\lim_n \mathbf{E}\xi_n \geqslant \lim_m \mathbf{E}\eta_m,$$

而由对称性, 有

$$\lim_n \mathbf{E}\xi_n \leqslant \lim_m \mathbf{E}\eta_m,$$

于是, (4) 式得证.

下面的注释往往是有益的.

注 1 非负随机变量的数学期望可以表示为:

$$\mathbf{E}\xi = \sup_{\{s \in S : s \leqslant \xi\}} \mathbf{E}s, \tag{6}$$

其中 $S = \{s\}$ 是非负简单随机变量的集合 (练习题 1).

这样, 对于非负简单随机变量, 定义了数学期望. 下面考虑一般情形.

设 ξ 是随机变量, $\xi^+ = \max(\xi, 0), \xi^- = -\min(\xi, 0)$.

定义 2 称随机变量 ξ 的数学期望 $\mathbf{E}\xi$ *存在或有定义*, 如果 $\mathbf{E}\xi^+$ 和 $\mathbf{E}\xi^-$ 中至少一个有限:

$$\min\{\mathbf{E}\xi^+, \mathbf{E}\xi^-\} < \infty.$$

这时, 随机变量 ξ 的数学期望*定义*为:

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^-.$$

数学期望 $\mathbf{E}\xi$ 又称为函数 ξ 对概率测度的*勒贝格积分*, (关于勒贝格积分的其他定义方法, 见第 11 小节.)

定义 3 称随机变量 ξ 的数学期望有限 (或者 ξ 可积), 如果 $\mathbf{E}\xi^+ < \infty$ 和 $\mathbf{E}\xi^- < \infty$.

因为 $|\xi| = \xi^+ + \xi^-$, 所以 $\mathbf{E}\xi$ 有限等价于 $\mathbf{E}|\xi| < \infty$. (在此意义上按勒贝格可积具有 "绝对的" 特点.)

注 2 除数学期望 $\mathbf{E}\xi$ 外, 随机变量 ξ 的重要数字特征还有 $\mathbf{E}\xi^r$ (如果它存在) 和 $\mathbf{E}|\xi|^r, r > 0$, 并相应地称为随机变量 ξ 的 $r-$ 阶矩和 $r-$ 阶绝对矩.

注 3 在上面给出的勒贝格积分 $\int_\Omega \xi \mathrm{d}\mathbf{P}$ 的定义, 曾假设: 测度 $\mathbf{P}$ 是概率测度 ($\mathbf{P}(\Omega) = 1$), 而 $\mathscr{F}-$ 可测函数 (随机变量) ξ 在 $\mathbb{R} = (-\infty, \infty)$ 中取值. 现在假设测度 μ 是定义在可测空间 $(\Omega, \mathscr{F})$ 上、且可能取 $+\infty$ 为值的任意测度, 而 $\xi = \xi(\omega)$ 是在 $\overline{\mathbb{R}} = [-\infty, \infty]$ 上取值的 $\mathscr{F}-$ 可测函数 (广义随机变量). 这时, 勒贝格积分

$$\int_\Omega \xi(\omega)\mu(\mathrm{d}\omega)$$

用同样的方法定义: 首先, 对非负简单函数 ξ (按 (2) 式将 $\mathbf{P}$ 换成 μ), 然后, 对任意非负简单函数 ξ, 一般按公式

$$\int_\Omega \xi(\omega)\mu(\mathrm{d}\omega) = \int_\Omega \xi^+ \mu(\mathrm{d}\omega) - \int_\Omega \xi^- \mu(\mathrm{d}\omega)$$

定义, 只要不出现 $\infty - \infty$ 的不定式.

对于数学分析, $(\Omega, \mathscr{F}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 且 μ 是勒贝格测度 λ 的情形特别重要. 这时, 将积分

$$\int_\mathbb{R} \xi(x)\lambda(\mathrm{d}x)$$

记作

$$\int_\mathbb{R} \xi(x)\mathrm{d}x, \text{ 或 } (\mathrm{L})\int_{-\infty}^{\infty} \xi(x)\mathrm{d}x,$$

以强调它与黎曼积分

$$(\mathrm{R})\int_{-\infty}^{\infty} \xi(x)\mathrm{d}x$$

的区别. 假如 (勒贝格 – 斯蒂尔切斯) 测度对应于某广义分布函数 $G = G(x)$, 则积分

$$\int_\mathbb{R} \xi(x)\mu(\mathrm{d}x)$$

也称做**勒贝格 – 斯蒂尔切斯积分**, 并记为

$$(\mathrm{L-S})\int_\mathbb{R} \xi(x)G(\mathrm{d}x),$$

以示其与相应的黎曼 – 斯蒂尔切斯积分 (见下面第 11 小节)

$$(\mathrm{R-S})\int_{-\infty}^{\infty} \xi(x)G(\mathrm{d}x)$$

的区别.

由下面性质 D, 可见假如定义了 $\mathbf{E}\xi$, 则对于任何 $A \in \mathscr{F}$, 也就定义了数学期望 $\mathbf{E}(\xi I_A)$, 亦常使用下列记号:

$$\mathbf{E}(\xi; I_A) \text{ 或 } \int_\Omega \xi I_A \mathrm{d}\mathbf{P}, \int_A \xi \mathrm{d}\mathbf{P},$$

其中最后一个积分习惯上称做 "ξ 对测度 $\mathbf{P}$ 在集合 A 上的勒贝格积分".

类似, 对于任意测度 μ, 下面两种表达式等价:

$$\int_\Omega \xi I_A \mathrm{d}\mu, \quad \int_A \xi \mathrm{d}\mu.$$

特别, 如果 μ 是 $n-$ 维勒贝格 $-$ 斯蒂尔切斯测度, $A = (a_1, b_1] \times \cdots \times (a_n, b_n]$, 则可以将 $\int_A \xi \mathrm{d}\mu$ 写成

$$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \xi(x_1, \cdots, x_n)\mu(\mathrm{d}x_1 \cdots \mathrm{d}x_n).$$

如果 μ 是勒贝格测度, 则将 $\mu(\mathrm{d}x_1 \cdots \mathrm{d}x_n)$ 简单写成 $\mathrm{d}x_1 \cdots \mathrm{d}x_n$.

3. 随机变量 ξ 的数学期望 $\mathbf{E}\xi$ 的性质

A. 若 c 是常数, 且 $\mathbf{E}\xi$ 存在, 则 $\mathbf{E}(c\xi)$ 也存在, 并且

$$\mathbf{E}(c\xi) = c\mathbf{E}\xi.$$

B. 若 $\xi \leqslant \eta$, 则

$$\mathbf{E}\xi \leqslant \mathbf{E}\eta,$$

即, 若 $-\infty < \mathbf{E}\xi$, 则

$$-\infty < \mathbf{E}\eta \text{ 且 } \mathbf{E}\xi \leqslant \mathbf{E}\eta,$$

或若 $\mathbf{E}\eta < \infty$, 则

$$\mathbf{E}\xi < \infty \text{ 且 } \mathbf{E}\xi \leqslant \mathbf{E}\eta.$$

C. 若 $\mathbf{E}\xi$ 存在, 则

$$|\mathbf{E}\xi| \leqslant \mathbf{E}|\xi|.$$

D. 若 $\mathbf{E}\xi$ 存在, 则对于每个 $A \in \mathscr{F}$, 数学期望 $\mathbf{E}(\xi I_A)$ 存在; 若 $\mathbf{E}\xi$ 有限, 则 $\mathbf{E}(\xi I_A)$ 也有限.

E. 若 ξ 和 η 是非负随机变量, 或满足 $\mathbf{E}|\xi| < \infty, \mathbf{E}|\eta| < \infty$, 则

$$\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta.$$

(关于这一性质的推广见练习题 2)

证明　性质 A ~ E 的证明.

(a) 对于简单随机变量, 命题显然. 设 $c \geqslant 0; \xi \geqslant 0, \xi_n \uparrow \xi$, 其中 ξ_n 是简单随机变量. 那么, 由于 $c\xi_n \uparrow c\xi$, 可见

$$\mathbf{E}(c\xi) = \lim_n \mathbf{E}(c\xi_n) = c \lim_n \mathbf{E}\xi_n = c\mathbf{E}\xi.$$

在一般情形下, 需要利用表示 $\xi = \xi^+ - \xi^-$, 并注意到, 对于 $c \geqslant 0, (c\xi)^+ = c\xi^+, (c\xi)^- = c\xi^-$, 而对于 $c < 0, (c\xi)^+ = -c\xi^-, (c\xi)^- = -c\xi^+$.

(b) 若 $0 \leqslant \xi \leqslant \eta$, 则 $\mathbf{E}\xi$ 和 $\mathbf{E}\eta$ 存在且 $\mathbf{E}\xi \leqslant \mathbf{E}\eta$, 由此 (6) 式得证. 现在设 $\mathbf{E}\xi > -\infty$, 则 $\mathbf{E}\xi^- < \infty$. 若 $\xi \leqslant \eta$, 则 $\xi^+ \leqslant \eta^+$ 和 $\xi^- \geqslant \eta^-$. 因此 $\mathbf{E}\eta^- \leqslant \mathbf{E}\xi^- < \infty$, 从而 $\mathbf{E}\eta$ 存在, 且

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^- \leqslant \mathbf{E}\eta^+ - \mathbf{E}\eta^- = \mathbf{E}\eta.$$

类似地考虑 $\mathrm{E}\eta < \infty$ 的情形.

(c) 由于 $-|\xi| \leqslant \xi \leqslant |\xi|$, 则由性质 **A** 和 **B**, 可见

$$-\mathbf{E}|\xi| \leqslant \mathbf{E}\xi \leqslant \mathbf{E}|\xi|,$$

即 $|\mathbf{E}\xi| \leqslant \mathbf{E}|\xi|$.

(d) 由性质 **B**, 以及

$$(\xi I_A)^+ = \xi^+ I_A \leqslant \xi^+, \ (\xi I_A)^- = \xi^- I_A \leqslant \xi^-,$$

可见 $\mathbf{E}(\xi I_A)$ 也有限.

(e) 设 $\xi_n \geqslant 0, \eta_n \geqslant 0$, 而 $\{\xi_n\}$ 和 $\{\eta_n\}$ 是简单函数序列, 且 $\xi_n \uparrow \xi$ 和 $\eta_n \uparrow \eta$. 那么,

$$\mathbf{E}(\xi_n + \eta_n) = \mathbf{E}\xi_n + \mathbf{E}\eta_n, \text{且}$$
$$\mathbf{E}(\xi_n + \eta_n) \uparrow \mathbf{E}(\xi + \eta), \mathbf{E}\xi_n \uparrow \mathbf{E}\xi, \mathbf{E}\eta_n \uparrow \mathbf{E}\eta,$$

从而 $\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta$. 对于 $\mathbf{E}|\xi| \leqslant \infty, \mathbf{E}|\eta| < \infty$ 的情形, 如果利用

$$\xi = \xi^+ - \xi^-, \eta = \eta^+ - \eta^-,$$
$$\xi^+ \leqslant |\xi|, \xi^- \leqslant |\xi|, \eta^+ \leqslant |\eta|, \eta^- \leqslant |\eta|,$$

则可以将其归结为讨论过的情形.

数学期望的进一步性质 数学期望的下一组性质 **F** $\sim$ **J**, 涉及 "**P**– 几乎必然" 的概念. 称某一性质 "**P**– 几乎必然" 成立, 如果存在概率为 0 的集合 $\mathscr{N} \in \mathscr{F}$, $\mathbf{P}(\mathscr{N}) = 0$, 使该性质对于每一点 $\omega \in \Omega - \mathscr{N}$ 成立. 与 "**P**– 几乎必然" (**P**–a.c.) 的概念同时, 常使用 "**P**– 几乎处处" (**P**–a. e.) 的概念, 或者简称 "几乎必然" (a.c.), 相应地简称 "几乎处处" (a.e.).

F. 若 $\xi = 0$ (a.c.),则$\mathbf{E}\xi = 0$.

事实上, 如果 ξ 是简单随机变量, $\xi = \sum_{k=1}^{n} x_k I_{A_k}(\omega), x_k \neq 0$, 则根据条件 $\mathbf{P}(A_k) = 0$, 因此 $\mathbf{E}\xi = 0$. 如果 $\xi > 0$ 且 $0 \leqslant s \leqslant \xi$, 其中 s 是简单随机变量, 故 $s = 0$ (a. c.), 从而 $\mathbf{E}s = 0$ 且

$$\mathbf{E}\xi = \sup_{\{s \in S: s \leqslant \xi\}} \mathbf{E}s = 0.$$

由于 $\xi = \xi^+ - \xi^-$, 且 $\xi^+ \leqslant |\xi|, \xi^- \leqslant |\xi|$, 而 $|\xi| = 0$ (a.c.), 所以, 通过极限过度, 可以将一般情形归结为 ξ 是简单随机变量的情形.

G. *若* $\xi = \eta$ (a.c.), *且* $\mathbf{E}|\xi| < \infty$, *则* $\mathbf{E}|\eta| < \infty$ *且* $\mathbf{E}\xi = \mathbf{E}\eta$ (*亦见练习题* 3).

实际上, 设 $\mathscr{N} = \{\omega : \xi \neq \eta\}$, 则 $\mathbf{P}(\mathscr{N}) = 0$ 且

$$\xi = \xi I_{\mathscr{N}} + \xi I_{\overline{\mathscr{N}}}, \quad \eta = \eta I_{\mathscr{N}} + \eta I_{\overline{\mathscr{N}}}.$$

由性质 **E** 和 **F**, 得

$$\mathbf{E}\xi = \mathbf{E}\xi I_{\mathscr{N}} + \mathbf{E}\xi I_{\overline{\mathscr{N}}} = \mathbf{E}\xi I_{\overline{\mathscr{N}}} = \mathbf{E}\eta I_{\overline{\mathscr{N}}}.$$

由于 $\mathbf{E}\eta I_{\mathscr{N}} = 0$, 故根据性质 **E**, 有 $\mathbf{E}\xi = \mathbf{E}\eta I_{\mathscr{N}} + \mathbf{E}\eta I_{\overline{\mathscr{N}}} = \mathbf{E}\eta$.

H. *若* $\xi \geqslant 0$ *且* $\mathbf{E}\xi = 0$, *则* $\xi = 0$ (a.c.).

为证明, 记 $A = \{\omega : \xi(\omega) > 0\}, A_n = \{\omega : \xi(\omega) \geqslant 1/n\}$. 显然, $A_n \uparrow A, 0 \leqslant \xi I_{A_n} \leqslant \xi I_A$. 因此由性质 **B** 有

$$0 \leqslant \mathbf{E}\xi I_{A_n} \leqslant \mathbf{E}\xi = 0.$$

从而

$$0 = \mathbf{E}\xi I_{A_n} \geqslant \frac{1}{n}\mathbf{P}(A_n),$$

即对于一切 $n \geqslant 1, \mathbf{P}(A_n) = 0$. 由于 $\mathbf{P}(A) = \lim \mathbf{P}(A_n)$, 故 $\mathbf{P}(A) = 0$.

I. *设* ξ *和* η *满足* $\mathbf{E}|\xi| < \infty, \mathbf{E}|\eta| < \infty$ *且对于一切* $A \in \mathscr{F}$, *有* $\mathbf{E}(\xi I_A) \leqslant \mathbf{E}(\eta I_A)$, *证明* $\xi \leqslant \eta$ (a.c.).

实际上, 设 $B = \{\omega : \xi(\omega) > \eta(\omega)\}$, 则 $\mathbf{E}(\eta I_B) \leqslant \mathbf{E}(\xi I_B) \leqslant \mathbf{E}(\eta I_B)$, 即 $\mathbf{E}(\xi I_B) = \mathbf{E}(\eta I_B)$. 由性质 **E**, 有 $\mathbf{E}((\xi - \eta)I_B) = 0$; 而由性质 **H**, 有 $(\xi - \eta)I_B = 0$ (a.c.), 因此 $\mathbf{P}(B) = 0$.

J. *设* ξ *是广义随机变量, 且* $\mathbf{E}|\xi| < \infty$. *证明* $|\xi| < \infty$ (a.c.).

事实上, 假设 $A = \{\omega : |\xi(\omega)| = \infty\}$, 且 $\mathbf{P}(A) > 0$, 则 $\mathbf{E}|\xi| \geqslant \mathbf{E}(|\xi| I_A) = \infty \times \mathbf{P}(A) = \infty$, 而这与假设 $\mathbf{E}|\xi| < \infty$ 矛盾 (亦见练习题 4).

4. 数学期望的极限定理 在这一小节讨论关于在数学期望 (勒贝格积分) 号下的极限过程.

定理 1 (单调收敛性) *设* $\eta, \xi, \xi_1, \xi_2, \cdots$ *是随机变量序列.*

a) *如果对于一切* $n \geqslant 1, \xi_n \geqslant \eta$, *有* $\mathbf{E}\eta > -\infty$, *且* $\xi_n \uparrow \xi$, *则*

$$\mathbf{E}\xi_n \uparrow \mathbf{E}\xi.$$

b) 如果对于一切 $n \geqslant 1, \xi_n \leqslant \eta$, 有 $\mathbf{E}\eta < \infty$, 且 $\xi_n \downarrow \xi$, 则

$$\mathbf{E}\xi_n \downarrow \mathbf{E}\xi.$$

证明 a) 首先假设 $\eta \geqslant 0$. 假设对于每一个 $k \geqslant 1$, 简单函数序列 $\{\xi_k^{(n)}\}_{n\geqslant 1}$ 当 $n \to \infty$ 时 $\xi_k^{(n)} \uparrow \xi_k$. 记 $\zeta^{(n)} = \max_{1\leqslant k\leqslant n} \xi_k^{(n)}$. 那么

$$\zeta^{(n-1)} \leqslant \zeta^{(n)} = \max_{1\leqslant k\leqslant n} \xi_k^{(n)} \leqslant \max_{1\leqslant k\leqslant n} \xi_k = \xi_n.$$

设 $\zeta = \lim_n \zeta^{(n)}$. 因为对于 $1 \leqslant k \leqslant n$,

$$\xi_k^{(n)} \leqslant \zeta^{(n)} \leqslant \xi_n.$$

那么, 若当 $n \to \infty$ 时取极限, 则对于任意 $k \geqslant 1$, 有

$$\xi_k \leqslant \zeta \leqslant \xi,$$

因此 $\xi = \zeta$.

由于 $\zeta^{(n)}$ 是简单随机变量且 $\zeta^{(n)} \uparrow \zeta$, 则

$$\mathbf{E}\xi = \mathbf{E}\zeta = \lim \mathbf{E}\zeta^{(n)} \leqslant \lim \mathbf{E}\xi_n.$$

另一方面, 显然, 由于 $\xi_n \leqslant \xi_{n+1} \leqslant \xi$, 则

$$\lim \mathbf{E}\xi_n \leqslant \mathbf{E}\xi.$$

从而 $\lim \mathbf{E}\xi_n = \mathbf{E}\xi$.

现在假设 η 是任意随机变量, 且 $\mathbf{E}\eta > -\infty$.

如果 $\mathbf{E}\eta = \infty$, 则由于性质 **B**, 可见 $\mathbf{E}\xi_n = \mathbf{E}\xi = \infty$, 从而命题得证. 设 $\mathbf{E}\eta < \infty$, 则考虑到关于的假设 $\mathbf{E}\eta > -\infty$, 得 $\mathbf{E}|\eta| < \infty$. 显然, 对于一切 $\omega \in \Omega, 0 \leqslant \xi_n - \eta \uparrow \xi - \eta$. 从而, 根据上面已证明的结果, 有 $\mathbf{E}(\xi_n - \eta) \uparrow \mathbf{E}(\xi - \eta)$, 因此 (由性质 **E**) 和练习题 2)

$$\mathbf{E}\xi_n - \mathbf{E}\eta \uparrow \mathbf{E}\xi - \mathbf{E}\eta.$$

因为 $\mathbf{E}|\eta| < \infty$, 所以当 $n \to \infty$ 时 $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi$.

如果将原变量加上负号, 则可以由命题 a) 得到命题 b). □

系 *设 $\{\eta_n\}_{n\geqslant 1}$ 是非负随机变量序列, 那么*

$$\mathbf{E}\sum_{n=1}^{\infty} \eta_n = \sum_{n=1}^{\infty} \mathbf{E}\eta_n.$$

证明 由数学期望的性质 **E** (亦见练习题 2), 单调收敛定理, 以及当 $k \to \infty$ 时

$$\sum_{n=1}^{k} \eta_n \uparrow \sum_{n=1}^{\infty} \eta_n,$$

即可得到. □

定理 2 (法图 [P. Fatou] 引理) 设 $\eta, \xi_1, \xi_2, \cdots$ 是随机变量.

a) 如果对于一切 $n \geqslant 1, \xi_n \geqslant \eta$, 且 $\mathbf{E}\eta > -\infty$, 则

$$\mathbf{E}\underline{\lim}\xi_n \leqslant \underline{\lim}\mathbf{E}\xi_n.$$

b) 如果对于一切 $n \geqslant 1, \xi_n \leqslant \eta$, 且 $\mathbf{E}\eta < \infty$, 则

$$\overline{\lim}\mathbf{E}\xi_n \leqslant \mathbf{E}\overline{\lim}\xi_n.$$

c) 如果对于一切 $n \geqslant 1, |\xi_n| \leqslant \eta$, 且 $\mathbf{E}\eta < \infty$, 则

$$\mathbf{E}\underline{\lim}\xi_n \leqslant \underline{\lim}\mathbf{E}\xi_n \leqslant \overline{\lim}\mathbf{E}\xi_n \leqslant \mathbf{E}\overline{\lim}\xi_n. \tag{7}$$

证明 设 $\zeta_n = \inf_{m \geqslant n} \xi_m$, 则

$$\underline{\lim}\xi_n = \lim_n \inf_{m \geqslant n} \xi_m = \lim_n \zeta_n.$$

明显地 $\zeta_n \uparrow \underline{\lim}\xi_n$, 且对于一切 $n \geqslant 1$, 有 $\zeta_n \geqslant \eta$. 那么, 有定理 1, 有

$$\mathbf{E}\underline{\lim_n}\xi_n = \mathbf{E}\lim_n \zeta_n = \lim_n \mathbf{E}\zeta_n = \underline{\lim_n}\mathbf{E}\zeta_n \leqslant \mathbf{E}\underline{\lim_n}\xi_n.$$

从而命题 a) 得证. 由 a) 可得 b). 由 a) 和 b) 可得 c). □

定理 3 (勒贝格控制收敛定理) 设随机变量 $\eta, \xi, \xi_1, \xi_2, \cdots$ 满足条件: $|\xi_n| \leqslant \eta$, $\mathbf{E}\eta < \infty$, 且 $\xi_n \to \xi$ (a.c.), 那么, $\mathbf{E}|\xi| < \infty$,

$$\mathbf{E}|\xi_n| \to \mathbf{E}\xi, n \to \infty, \tag{8}$$

且

$$\mathbf{E}|\xi_n - \xi| \to 0, n \to \infty. \tag{9}$$

证明 根据假设 $\underline{\lim}\xi_n = \overline{\lim}\xi_n = \xi$ (a.c.). 因此, 由性质 **G** 和法图引理 (命题 c))

$$\mathbf{E}\xi = \mathbf{E}\underline{\lim_n}\xi_n \leqslant \underline{\lim_n}\mathbf{E}\xi_n = \overline{\lim_n}\mathbf{E}\xi_n = \mathbf{E}\overline{\lim_n}\xi_n = \mathbf{E}\xi,$$

因而 (8) 式得证. 显然, $|\xi| \leqslant \eta$. 所以 $\mathbf{E}|\xi| < \infty$.

只需注意到 $|\xi_n - \xi| \leqslant 2\eta$, 即可证明命题 (9).

系 设随机变量 $\eta, \xi, \xi_1, \xi_2, \cdots$ 满足条件: $|\xi_n| \leqslant \eta, \xi_n \to \xi$ (a.c.), 且对于某个 $p > 0$, $\mathbf{E}\eta^p < \infty$, 则 $\mathbf{E}|\xi|^p < \infty$, 且当 $n \to \infty$ 时 $\mathbf{E}|\xi_n - \xi|^p \to 0$.

为证明此系, 只需注意到, $|\xi| \leqslant \eta$ 和

$$|\xi_n - \xi|^p \leqslant (|\xi_n| + |\xi|)^p \leqslant (2\eta)^p.$$

在法图引理和控制收敛定理中, 保障公式 (7) ~ (9) 成立的条件 "$|\xi_n| \leqslant \eta, \mathbf{E}\eta < \infty$", 可以略为减弱. 为表述相应的结果 (定理 4), 现在给出如下定义.

定义 4 随机变量族 $\{\xi_n\}_{n\geqslant 1}$ 称做 (对测度 $\mathbf{P}$) 一致可积的, 如果

$$\sup_n \int_{\{|\xi_n|>c\}} |\xi_n|\mathbf{P}(\mathrm{d}\omega) \to 0, c \to \infty, \tag{10}$$

或 (在其他记号下)

$$\sup_n \mathbf{E}[|\xi_n|I_{\{|\xi_n|>c\}}] \to 0, c \to \infty, \tag{11}$$

显然, 如果随机变量 $\xi_n, n \geqslant 1$ 满足条件 $|\xi_n| \leqslant \eta, \mathbf{E}\eta < \infty$, 则随机变量族 $\{\xi_n\}_{n\geqslant 1}$ 是一致可积的.

定理 4 *设随机变量族* $\{\xi_n\}_{n\geqslant 1}$ *是一致可积的.*

a) *那么*

$$\mathbf{E}\varliminf_n \xi_n \leqslant \varliminf_n \mathbf{E}\xi_n \leqslant \varlimsup_n \mathbf{E}\xi_n = \mathbf{E}\varlimsup_n \xi_n.$$

b) *此外, 如果* $\xi_n \to \xi$ (a.c.), *则随机变量* ξ *可积, 且*

$$\mathbf{E}\xi_n \to \mathbf{E}\xi, \quad n \to \infty,$$
$$\mathbf{E}|\xi_n - \xi| \to 0, \quad n \to \infty.$$

证明 a) 对于任意 $c > 0$, 有

$$\mathbf{E}\xi_n = \mathbf{E}\xi_n I_{\{\xi_n<-c\}} + \mathbf{E}\xi_n I_{\{\xi_n\geqslant -c\}}. \tag{12}$$

由于一致可积性, 对于任意 $\varepsilon > 0$, 可以选择充分大的 c, 使

$$\sup_n |\mathbf{E}\xi_n I_{\{\xi_n<-c\}}| < \varepsilon. \tag{13}$$

由法图引理, 可见

$$\varliminf \mathbf{E}\xi_n I_{\{\xi_n\geqslant -c\}} \geqslant \mathbf{E}\varliminf \xi_n I_{\{\xi_n\geqslant -c\}}.$$

因为 $\xi_n I_{\{\xi_n\geqslant -c\}} \geqslant \xi_n$, 所以

$$\varliminf \mathbf{E}\xi_n I_{\{\xi_n\geqslant -c\}} \geqslant \mathbf{E}\varliminf \xi_n. \tag{14}$$

由 (12)~(14) 式可得

$$\varliminf \mathbf{E}\xi_n \geqslant \mathbf{E}\varliminf \xi_n - \varepsilon.$$

由于 $\varepsilon > 0$ 是任意的, 可见 $\varliminf \mathbf{E}\xi_n \geqslant \mathbf{E}\varliminf \xi_n$.

类似地, 可以证明: $\varlimsup \mathbf{E}\xi_n \leqslant \mathbf{E}\varlimsup \xi_n$.

至于命题 b), 则其证明与定理 3 相应命题的证明相同. □

下面的定理, 更完整地揭示了控制收敛定理的意义, 并给出了在数学期望号下取极限的充分和必要条件.

定理 5　设 $0 \leqslant \xi_n \to \xi(\mathbf{P}\text{-a.c.})$, 而 $\mathbf{E}\xi_n < \infty$. 那么, $\mathbf{E}\xi_n \to \mathbf{E}\xi < \infty$ 的充分和必要条件是: 随机变量族 $\{\xi_n\}_{n\geqslant 1}$ 一致可积.

证明　由定理 4 的命题 b) 可见充分性成立. 为证明必要性, 考虑 (有限或可数) 集合 $A = \{a : \mathbf{P}\{\xi = a\} > 0\}$. 那么, 对于每一个 $a \notin A, \xi_n I_{\{\xi_n<a\}} \to \xi I_{\{\xi<a\}}$. 并且随机变量族 $\{\xi_n I_{\{\xi_n<a\}}\}_{n\geqslant 1}$ 一致可积. 因此, 由于 "充分性" 知, 对于每一个 $a \notin A, \mathbf{E}\xi_n I_{\{\xi_n<a\}} \to \mathbf{E}\xi I_{\{\xi<a\}}$, 因而当 $n \to \infty$ 时

$$\mathbf{E}\xi_n I_{\{\xi_n\geqslant a\}} \to \mathbf{E}\xi I_{\{\xi\geqslant a\}}, a \notin A. \tag{15}$$

固定 $\varepsilon > 0$, 选取充分大的 $a_0 \notin A$, 使 $\mathbf{E}\xi I_{\{\xi\geqslant a_0\}} < \varepsilon/2$; 然后选取 N_0, 使对于一切 $n \geqslant N_0$, 有

$$\mathbf{E}\xi_n I_{\{\xi_n\geqslant a_0\}} \leqslant \mathbf{E}\xi I_{\{\xi_n\geqslant a_0\}} + \frac{\varepsilon}{2},$$

因而 $\mathbf{E}\xi_n I_{\{\xi_n\geqslant a_0\}} \leqslant \varepsilon$. 最后, 选取充分大的 $a_1 \geqslant a_0$, 使对于一切 $n \leqslant N_0 \mathbf{E}\xi_n I_{\{\xi_n\geqslant a_1\}} \leqslant \varepsilon$. 那么,

$$\sup_n \mathbf{E}\xi_n I_{\{\xi_n\geqslant a_1\}} \leqslant \varepsilon,$$

于是, $\{\xi_n\}_{n\geqslant 1}$ 的一致可积性得证. □

5. 一致可积的准则　首先指出, 如果随机变量族 $\{\xi_n\}_{n\geqslant 1}$ 一致可积, 则

$$\sup_n \mathbf{E}|\xi_n| < \infty. \tag{16}$$

实际上, 对于固定的 $\varepsilon > 0$ 和充分大的 $c > 0$,

$$\begin{aligned}\sup_n \mathbf{E}|\xi_n| &= \sup_n [\mathbf{E}|\xi_n| I_{\{|\xi_n|\geqslant c\}} + \mathbf{E}|\xi_n| I_{\{|\xi_n|<c\}}] \\ &\leqslant \sup_n \mathbf{E}|\xi_n| I_{\{|\xi_n|\geqslant c\}} + \sup_n \mathbf{E}|\xi_n| I_{\{|\xi_n|<c\}} \leqslant \varepsilon + c.\end{aligned}$$

因此 (16) 式得证.

结果表明, 条件 (16) 连同所谓 "一致连续性" 条件, 是一致可积性的充分和必要条件.

引理 2　随机变量族 $\{\xi_n\}_{n\geqslant 1}$ 一致可积的充分和必要条件, 是 $\mathbf{E}|\xi_n| (n \geqslant 1)$ 一致有界 (即 (16) 式成立) 和 $\mathbf{E}\{|\xi_n| I_A\} (n \geqslant 1)$ 一致连续 (即当 $\mathbf{P}(A) \to 0$ 时 $\sup_n \mathbf{E}\{|\xi_n| I_A\} \to 0$).

证明　*必要性*. 上面已经验证过条件 (16). 其次, 有

$$\begin{aligned}\mathbf{E}\{|\xi_n| I_A\} &= \mathbf{E}\{|\xi_n| I_{A\cap\{|\xi_n|\geqslant c\}}\} + \mathbf{E}\{|\xi_n| I_{A\cap\{|\xi_n|<c\}}\} \\ &\leqslant \mathbf{E}\{|\xi_n| I_{\{|\xi_n|\geqslant c\}}\} + c\mathrm{P}(A).\end{aligned} \tag{17}$$

选择 c 充分大, 使

$$\sup_n \mathbf{E}\{|\xi_n| I_{\{|\xi_n|\geqslant c\}}\} \leqslant \frac{\varepsilon}{2},$$

那么, 如果 $\mathbf{P}(A) \leqslant \varepsilon/2c$, 则由 (17) 式, 可见

$$\sup_n \mathbf{E}\{|\xi_n| I_A\} \leqslant \varepsilon.$$

因此必要性得证.

充分性. 设 $\varepsilon > 0, \delta > 0$, 由条件 $\mathbf{P}(A) < \delta$, 对 n 一致, 有 $\mathbf{E}\{|\xi_n| I_A\} \leqslant \varepsilon$. 因为对于任意 $c > 0$,

$$\mathbf{E}|\xi_n| \geqslant \mathbf{E}\{|\xi_n| I_{\{|\xi_n| \geqslant c\}}\} \geqslant c\mathbf{P}\{|\xi_n| \geqslant c\}$$

(对照切比雪夫不等式), 所以

$$\sup_n \mathbf{P}\{|\xi_n| \geqslant c\} \leqslant \frac{1}{c} \sup_n \mathbf{E}|\xi_n| \to 0, c \to \infty.$$

因此, 对于充分大的 c, 作为集合 A 可以取集合 $\{|\xi_n| \geqslant c\} (n \geqslant 1)$ 中的任意一个. 于是, 一致可积性得证. □

下面的引理给出了便于应用的一致可积性的充分条件.

引理 3 设 $\xi_1, \xi_2, \cdots$ 是可积随机变量序列, 而 $G = G(t)$ 是定义在 $[0, \infty)$ 上的非负增函数, 满足

$$\lim_{t \to \infty} \frac{G(t)}{t} = \infty, \tag{18}$$

$$\sup_n \mathbf{E}G(|\xi_n|) = \infty. \tag{19}$$

那么, 随机变量族 $\{\xi_n\}_{n \geqslant 1}$ 一致可积.

证明 设 $\varepsilon > 0$,

$$M = \sup_n \mathbf{E}G(|\xi_n|), a = \frac{M}{\varepsilon}.$$

取 c 充分大, 使对于 $t \geqslant c$, 有 $G(t)/t \geqslant a$. 那么, 关于 $n \geqslant 1$ 一致有

$$\mathbf{E}[|\xi_n| I_{\{|\xi_n| \geqslant c\}}] \leqslant \frac{1}{a} \mathbf{E}[G(|\xi_n|) I_{\{|\xi_n| \geqslant c\}}] \leqslant \frac{M}{a} = \varepsilon.$$ □

6. 独立随机变量之积的数学期望 如同第一章 §4 第 5 小节, 对于 ξ 和 η 是独立简单随机变量的情形, 证明了 $\mathbf{E}\xi\eta = \mathbf{E}\xi \times \mathbf{E}\eta$. 现在证明类似命题的一般情形 (亦见练习题 6).

定理 6 假设 ξ 和 η 是独立随机变量, 且 $\mathbf{E}|\xi| < \infty, \mathbf{E}|\eta| < \infty$. 那么, $\mathbf{E}|\xi\eta| < \infty$, 且

$$\mathbf{E}\xi\eta = \mathbf{E}\xi \times \mathbf{E}\eta. \tag{20}$$

证明 首先, 假设 $\xi \geqslant 0, \eta \geqslant 0$. 记

$$\xi_n = \sum_{k=0}^{\infty} \frac{k}{n} I_{\{\frac{k}{n} \leqslant \xi < \frac{k+1}{n}\}}, \eta_n = \sum_{k=0}^{\infty} \frac{k}{n} I_{\{\frac{k}{n} \leqslant \eta < \frac{k+1}{n}\}},$$

则 $\xi_n \leqslant \xi, |\xi_n - \xi| \leqslant 1/n$ 和 $\eta_n \leqslant \eta, |\eta_n - \eta| \leqslant 1/n$. 由于 $\mathbf{E}\xi < \infty, \mathbf{E}\eta < \infty$, 则根据勒贝格控制收敛定理,

$$\lim_n \mathbf{E}\xi_n = \mathbf{E}\xi, \quad \lim_n \mathbf{E}\eta_n = \mathbf{E}\eta.$$

其次, 由于 ξ 和 η 独立性, 当 $n \to \infty$ 时, 有

$$\begin{aligned}\mathbf{E}\xi_n\eta_n &= \sum_{k,l\geqslant 0} \frac{kl}{n^2}\mathbf{E}I_{\{\frac{k}{n}\leqslant\xi<\frac{k+1}{n}\}}I_{\{\frac{l}{n}\leqslant\eta<\frac{l+1}{n}\}} \\ &= \sum_{k,l\geqslant 0} \frac{kl}{n^2}\mathbf{E}I_{\{\frac{k}{n}\leqslant\xi<\frac{k+1}{n}\}} \times \mathbf{E}I_{\{\frac{l}{n}\leqslant\eta<\frac{l+1}{n}\}} = \mathbf{E}\xi_n \times \mathbf{E}\eta_n.\end{aligned}$$

因此, 当 $n \to \infty$ 时, 有

$$\begin{aligned}|\mathbf{E}\xi\eta - \mathbf{E}\xi_n\eta_n| &\leqslant \mathbf{E}|\xi\eta - \xi_n\eta_n| \leqslant \mathbf{E}[\xi|\eta - \eta_n|] + \mathbf{E}[\eta_n|\xi - \xi_n|] \\ &\leqslant \frac{1}{n}\mathbf{E}\xi + \frac{1}{n}\mathbf{E}\left(\eta + \frac{1}{n}\right) \to 0.\end{aligned}$$

于是,

$$\mathbf{E}\xi\eta = \lim_n \mathbf{E}\xi_n\eta_n = \lim_n \mathbf{E}\xi_n \times \lim_n \eta_n = \mathbf{E}\xi \times \mathbf{E}\eta,$$

并且 $\mathbf{E}\xi\eta < \infty$.

如果利用关系式

$$\xi = \xi^+ - \xi^-, \eta = \eta^+ - \eta^-, \xi\eta = \xi^+\eta^+ - \xi^-\eta^+ - \xi^+\eta^- + \xi^-\eta^-,$$

则一般情形可以归结为上面讨论的情形. □

7. 与数学期望有关的不等式 现在证明与数学期望有关的若干不等式, 这些不等式在概率论以及数学分析中都有系统的应用, 其中许多已经在初等概率论中介绍过 (见第一章 §4 和 §5).

切比雪夫不等式. *设 ξ 是非负随机变量, 则对于任意 $\varepsilon > 0$, 有*

$$\mathbf{P}\{\xi \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}\xi}{\varepsilon}. \tag{21}$$

证明 易见

$$\mathbf{E}\xi \geqslant \mathbf{E}[\xi I_{\{\xi\geqslant\varepsilon\}}] \geqslant \varepsilon\mathbf{E}I_{\{\xi\geqslant\varepsilon\}} = \varepsilon\mathbf{P}\{\xi \geqslant \varepsilon\},$$

由此立即得不等式 (21). 由不等式 (21), 可得切比雪夫不等式的如下等价形式: *如果是任意随机变量 ξ, 则*

$$\mathbf{P}\{\xi \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}\xi^2}{\varepsilon^2} \tag{22}$$

和

$$\mathbf{P}\{|\xi - \mathbf{E}\xi| \geqslant \varepsilon\} \leqslant \frac{\mathbf{D}\xi}{\varepsilon^2}, \tag{23}$$

其中 $\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2$ 是随机变量 ξ 的方差,

柯西 – 布尼亚科夫斯基不等式 假设对于随机变量 ξ 和 η,$\mathbf{E}\xi^2 < \infty, \mathbf{E}\eta^2 < \infty$. 那么, $\mathbf{E}|\xi\eta| < \infty$ 且

$$(\mathbf{E}|\xi\eta|)^2 \leqslant \mathbf{E}\xi^2 \times \mathbf{E}\eta^2. \tag{24}$$

证明 假设 $\mathbf{E}\xi^2 > 0, \mathbf{E}\eta^2 > 0$. 那么, 记

$$\widetilde{\xi} = \frac{\xi}{\sqrt{\mathbf{E}\xi^2}}, \quad \widetilde{\eta} = \frac{\eta}{\sqrt{\mathbf{E}\eta^2}},$$

则由 $2|\widetilde{\xi}\widetilde{\eta}| \leqslant \widetilde{\xi}^2 + \widetilde{\eta}^2$, 可见,

$$2\mathbf{E}|\widetilde{\xi}\widetilde{\eta}| \leqslant \mathbf{E}\widetilde{\xi}^2 + \mathbf{E}\widetilde{\eta}^2 = 2,$$

即 $\mathbf{E}|\widetilde{\xi}\widetilde{\eta}| \leqslant 1$, 由此得不等式 (24).

假如 $\mathbf{E}\xi^2 \equiv 0$, 则根据性质 **I**,$\xi = 0$ (a.c.), 而根据性质 **F**, $\mathbf{E}\xi\eta = 0$, 于是, 不等式 (24) 仍然成立.

延森 (I. L. Jencen) 不等式 设对于 $x \in \mathbb{R}, g = g(x)$ 是凹 (亦称下凸) 博雷尔函数, 且 $\mathbf{E}|\xi| < \infty$. 那么,

$$g(\mathbf{E}\xi) \leqslant \mathbf{E}g(\xi). \tag{25}$$

证明 由于 $g = g(x)$ 是凹函数, 故对于每一点 $x_0 \in \mathbb{R}$, 存在 $\lambda(x_0)$, 使对于一切 $x \in R$,

$$g(x) \geqslant g(x_0) + (x - x_0)\lambda(x_0). \tag{26}$$

设 $x = \xi$ 和 $x_0 = \mathbf{E}\xi$, 则由 (26) 式可见

$$g(\xi) \geqslant g(\mathbf{E}\xi) + (x - \mathbf{E}\xi)\lambda(\mathbf{E}\xi).$$

于是, 在上式两侧同求数学期望, 得 $\mathbf{E}g(\xi) \geqslant g(\mathbf{E}\xi)$. □

由延森不等式可以导出一系列有用的不等式. 下面举几个例子.

李亚普诺夫不等式 对于 $0 < s < t$, 有

$$(\mathbf{E}|\xi|^s)^{1/s} \leqslant (\mathbf{E}|\xi|^t)^{1/t}. \tag{27}$$

易见, 若记 $r = t/s, \eta = |\xi|^s$, 并在延森不等式中设 $g(x) = |x|^r$, 则得 $|\mathbf{E}\eta|^r \leqslant \mathbf{E}|\eta|^r$, 即

$$(\mathbf{E}|\xi|^s)^{t/s} \leqslant \mathbf{E}|\xi|^t,$$

由此立即得不等式 (27).

由李亚普诺夫不等式, 得如下绝对矩之间的一系列不等式:

$$\mathbf{E}|\xi| \leqslant (\mathbf{E}|\xi|^2)^{1/2} \leqslant \cdots \leqslant (\mathbf{E}|\xi|^n)^{1/n}. \tag{28}$$

赫尔德 (O. L. Hölder) 不等式　设 $1 < p < \infty, 1 < q < \infty$ 且 $1/p + 1/q = 1$. 如果 $\mathbf{E}|\xi|^p < \infty, \mathbf{E}|\eta|^q < \infty$, 则 $\mathbf{E}|\xi\eta| < \infty$ 且

$$\mathbf{E}|\xi\eta| \leqslant (\mathbf{E}|\xi|^p)^{1/p} (\mathbf{E}|\eta|^q)^{1/q}. \tag{29}$$

证明　如果 $\mathbf{E}|\xi|^p = 0$ 或 $\mathbf{E}|\eta|^q = 0$, 则立即得不等式 (29), 就像柯西 – 李亚普诺夫不等式的情形一样 (当 $p = q = 2$ 时, 赫尔德不等式就是柯西 – 李亚普诺夫不等式).

现在设 $\mathbf{E}|\xi|^p > 0, \mathbf{E}|\eta|^q > 0$. 记

$$\widetilde{\xi} = \frac{|\xi|}{(\mathbf{E}|\xi|^p)^{1/p}}, \quad \widetilde{\eta} = \frac{|\eta|}{(\mathbf{E}|\eta|^q)^{1/q}}.$$

不难证明不等式

$$x^a y^b \leqslant ax + by, \tag{30}$$

其中 $x > 0, y > 0, a > 0, b > 0, a + b = 1$. 实际上, 由对数函数的凸性, 可见

$$\ln(ax + by) \geqslant a\ln\ x + b\ln\ y = \ln\ x^a y^b;$$

经对数还原, 立即得 (30) 式.

在 (30) 式中设 $x = \widetilde{\xi}^p, y = \widetilde{\eta}^q, a = 1/p, b = 1/q$, 得

$$\widetilde{\xi}\widetilde{\eta} \leqslant \frac{1}{p}\widetilde{\xi}^p + \frac{1}{q}\widetilde{\eta}^q,$$

$$\mathbf{E}\widetilde{\xi}\widetilde{\eta} \leqslant \frac{1}{p}\mathbf{E}\widetilde{\xi}^p + \frac{1}{q}\mathbf{E}\widetilde{\eta}^q = \frac{1}{p} + \frac{1}{q} = 1.$$

于是, (29) 式得证. □

闵可夫斯基 (Г. Минковский) 不等式　如果 $\mathbf{E}|\xi|^p < \infty, \mathbf{E}|\eta|^p < \infty, 1 \leqslant p < \infty$, 则 $\mathbf{E}|\xi + \eta|^p < \infty$, 且

$$(\mathbf{E}|\xi + \eta|^p)^{1/p} \leqslant (\mathbf{E}|\xi|^p)^{1/p} + (\mathbf{E}|\eta|^p)^{1/p}. \tag{31}$$

证明　首先证明如下代数不等式: 如果 $a, b > 0$ 和 $p \geqslant 1$, 则

$$(a + b)^p \leqslant 2^{p-1}(a^p + b^p). \tag{32}$$

实际上, 考虑函数 $F(x) = (a + x)^p - 2^{p-1}(a^p + x^p)$. 那么

$$F'(x) = p(a + x)^{p-1} - 2^{p-1}px^{p-1},$$

而由于 $p \geqslant 1$, 可见对于 $x < a, F'(a) = 0, F'(x) > 0$, 而对于 $x > a, F'(x) < 0$. 因此,

$$F(b) \leqslant \max\ F(x) = F(a) = 0,$$

从而不等式 (32) 得证.

利用不等式 (32), 得

$$|\xi+\eta|^p \leqslant (|\xi|+|\eta|)^p \leqslant 2^{p-1}(|\xi|^p+|\eta|^p). \tag{33}$$

因此, 如果 $\mathbf{E}|\xi|^p<\infty, \mathbf{E}|\eta|^p<\infty$, 则 $\mathbf{E}|\xi+\eta|^p<\infty$.

如果 $p=1$, 则由 (33) 式可得 (31) 式.

现在假设 $p>1$. 取 $q>1$, 使 $1/p+1/q=1$, 则

$$|\xi+\eta|^p = |\xi+\eta||\xi+\eta|^{p-1} \leqslant |\xi||\xi+\eta|^{p-1}+|\eta||\xi+\eta|^{p-1}. \tag{34}$$

因为 $(p-1)q=p$, 所以

$$\mathbf{E}(|\xi+\eta|^{p-1})^q = \mathbf{E}|\xi+\eta|^p<\infty.$$

因此, 由赫尔德不等式, 有

$$\mathbf{E}(|\xi||\xi+\eta|^{p-1}) \leqslant (\mathbf{E}|\xi|^p)^{1/p}(\mathbf{E}|\xi+\eta|^{(p-1)q})^{1/q} = (\mathbf{E}|\xi|^p)^{1/p}(\mathbf{E}|\xi+\eta|^p)^{1/q}.$$

同样可得

$$\mathbf{E}(|\eta||\xi+\eta|^{p-1}) \leqslant (\mathbf{E}|\eta|^p)^{1/p}(\mathbf{E}|\xi+\eta|^p)^{1/q}.$$

从而, 由 (34) 式有

$$\mathbf{E}|\xi+\eta|^p \leqslant (\mathbf{E}|\xi+\eta|^p)^{1/q}\left[(\mathbf{E}|\xi|^p)^{1/p}+(\mathbf{E}|\eta|^p)^{1/p}\right]. \tag{35}$$

如果 $\mathbf{E}|\xi+\eta|^p=0$, 则欲证明的 (31) 式显然. 现在假设 $\mathbf{E}|\xi+\eta|^p>0$, 则由 (35) 式可见

$$(\mathbf{E}|\xi+\eta|^p)^{1-1/q} \leqslant \left[(\mathbf{E}|\xi|^p)^{1/p}+(\mathbf{E}|\eta|^p)^{1/p}\right].$$

于是, 由于 $1-1/q=1/p$, 得所要证明的 (31) 式. □

8. 拉东 – 尼科迪姆 (J. Radon-O. M. Nikodým) 定理 设随机变量 ξ 有数学期望 $\mathbf{E}\xi$. 那么, 由性质 D, 决定一集函数

$$\mathbf{Q}(A) \equiv \int_A \xi \mathrm{d}\mathbf{P}, A\in\mathscr{F}. \tag{36}$$

现在证明, 该函数是*可数 – 可加的*.

首先假设 ξ 是非负随机变量. 若 $A_1, A_2, \cdots$ 是 $\mathscr{F}$ 中的两两不相交集合, 记 $A=\sum A_n$, 则根据定理 1 的系, 有

$$\mathbf{Q}(A)=\mathbf{E}(\xi I_A)=\mathbf{E}\left(\xi I_{\sum A_n}\right)=\mathbf{E}(\sum \xi I_{A_n})=\sum \mathbf{E}(\xi I_{A_n})=\sum \mathbf{Q}(A_n).$$

如果 ξ 是有数学期望 $\mathbf{E}\xi$ 的任意随机变量, 则由下列关系式 (a)~(c), 可见函数 $\mathbf{Q}(A)$ 的可数 – 可加:

(a) 表现

$$\mathbf{Q}(A)=\mathbf{Q}^{+}(A)-\mathbf{Q}^{-}(A), \tag{37}$$

其中

$$\mathbf{Q}^{+}(A)=\int_A \xi^{+}\mathrm{d}\mathbf{P}, \mathbf{Q}^{-}(A)=\int_A \xi^{-}\mathrm{d}\mathbf{P};$$

(b) (上面已证明的) 非负随机变量的可数 - 可加性;

(c) $\min\{\mathbf{Q}^{+}(\Omega),\mathbf{Q}^{-}(\Omega)\}<\infty$.

这样, 如果 $\mathbf{E}\xi$ 存在, 则函数 $\mathbf{Q}=\mathbf{Q}(A)$ 是带符号的测度, 即可表示为 $\mathbf{Q}=\mathbf{Q}_1-\mathbf{Q}_2$ 的可数 – 可加集函数, 其中测度 Q_1 或 Q_2 至少有一个有限.

现在证明, 函数 $\mathbf{Q}=\mathbf{Q}(A)$ 具有如下重要性质: 关于测度 $\mathbf{P}$ 绝对连续:

$$\text{如果 } \mathbf{P}(A)=0, \text{ 则 } \mathbf{Q}(A)=0, \quad (A\in\mathscr{F}).$$

(该性质简记为 $\mathbf{Q}\ll\mathbf{P}$).

为证明这一事实, 只需考虑非负随机变量的情形. 如果

$$\xi=\sum_{k=1}^{n}x_k I_{A_k}$$

是简单非负随机变量, 且 $\mathbf{P}(A)=0$, 则

$$\mathbf{Q}(A)=\mathbf{E}(\xi I_A)=\sum_{k=1}^{n}x_k\mathbf{P}(A_k\cap A)=0.$$

如果 $\{\xi_n\}_{n\geqslant1}$ 是非负简单函数序列, 且 $\xi_n\uparrow\xi\geqslant0$, 则根据控制收敛定理, 有

$$\mathbf{Q}(A)=\mathbf{E}(\xi I_A)=\lim\mathbf{E}(\xi_n I_{A_n})=0,$$

因为对于任何 $n\geqslant1$ 及使 $\mathbf{P}(A)=0$ 的 $A, \mathbf{E}(\xi_n I_{A_n})=0$.

于是, 勒贝格积分

$$\mathbf{Q}(A)=\int_A \xi\mathrm{d}\mathbf{P},$$

作为集合 $A\in\mathscr{F}$ 的函数, 是关于测度 $\mathbf{P}(\mathbf{Q}\ll\mathbf{P})$ 绝对连续的带符号的测度. 非常值得注意的是逆命题也成立.

拉东 - 尼科迪姆定理 设 $(\Omega,\mathscr{F})$ 是可测空间,μ 是 $\sigma-$ 有限测度, 而 λ 是关于 μ 绝对连续的带符号的测度 (即 $\lambda=\lambda_1-\lambda_2$, 其中测度 λ_1 或 λ_2 至少有一个有限). 那么, 存在在 $\overline{\mathbb{R}}=[-\infty,\infty]$ 上取值的 $\mathscr{F}-$ 可测函数 $f=f(\omega)$, 使

$$\lambda(A)=\int_A f(\omega)\mu(\mathrm{d}\omega), \quad A\in\mathscr{F}. \tag{38}$$

函数 $f=f(\omega)$ 精确到 $\mu-$ 测度 0 唯一: 倘若 $h=h(\omega)$ 是另一 $\mathscr{F}-$ 可测函数, 使

$$\lambda(A)=\int_A h(\omega)\mu(\mathrm{d}\omega), \quad A\in\mathscr{F},$$

则 $\mu\{\omega: f(\omega) \neq h(\omega)\} = 0$.

如果 μ 是测度, 则 $f = f(\omega)$ 的值域为 $\overline{\mathbb{R}}_+ = [0, \infty]$.

式 (38) 中的函数 $f = f(\omega)$, 称做*拉东 – 尼科迪姆导数*, 或测度 λ 关于测度 μ 的*密度*, 并且记作

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} \quad 或 \quad \frac{\mathrm{d}\lambda}{\mathrm{d}\mu}(\omega).$$

这些导数的一系列性质, 将在下面的 §7 中第 8 小节介绍. 特别指出 §7 中的公式 (35), 在置换测度的情况下重新计算数学期望时常要用到.

具体地说, 设 $\mathbf{P}$ 和 $\widetilde{\mathbf{P}}$ 是两个概率测度,$\mathbf{E}$ 和 $\widetilde{\mathbf{E}}$ 是相应的数学期望. 假设测度 $\widetilde{\mathbf{P}}$ 关于测度 $\mathbf{P}$ 绝对连续 ($\widetilde{\mathbf{P}} \ll \mathbf{P}$). 那么, 对于一切非负随机变量 $\xi = \xi(\omega)$, 有如下 "*数学期望换算公式*":

$$\widetilde{\mathbf{E}}\xi = \mathbf{E}\left(\xi \frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\right). \tag{39}$$

在不假定随机变量 ξ 非负的情况下, *数学期望换算公式*仍然成立: 在随机变量 ξ 关于测度 $\widetilde{\mathbf{P}}$ 可积, 当且仅当随机变量 $\xi\dfrac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}$ 关于测度 $\mathbf{P}$ 可积; 这时 (39) 式成立.

公式 (39) 的证明并不复杂: 对于简单函数 ξ 它可以由导数 $\mathrm{d}\widetilde{\mathbf{P}}/\mathrm{d}\mathbf{P}$ 的定义得到; 对于非负函数 ξ, 首先, 由 §4 定理 1 的 b) 可见, 存在简单函数 $\xi_n \uparrow \xi (n \to \infty)$; 其次, 利用由 §4 中单调函数收敛的定理 1 的 a), 即可证明 (39) 式. 如果 ξ 是任意随机变量, 则根据 (39) 式, 有

$$\widetilde{\mathbf{E}}|\xi| = \mathbf{E}|\xi|\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}.$$

"ξ 对测度 $\widetilde{\mathbf{P}}$ 的可积性" 与 "$\xi(\mathrm{d}\widetilde{\mathbf{P}}/\mathrm{d}\mathbf{P})$ 对测度 $\mathbf{P}$ 的可积性" 等价. 至于公式 (39) 本身的证明, 只需要考虑表示 $\xi = \xi^+ - \xi^-$ 即可.

这里不加证明直接引用的拉东 – 尼科迪姆定理 (例如, 其证明可以参见 [70]), 将在 (§7) 建立条件数学期望时起关键作用.

9. 勒贝格积分中的变量替换 如果

$$\xi = \sum_{i=1}^{n} x_i I_{A_i}$$

是简单随机变量, $A_i = \{\omega : \xi = x_i\}$, 则

$$\mathbf{E}g(\xi) = \sum g(x_i)\mathbf{P}(A_i) = \sum g(x_i)\Delta F_\xi(x_i).$$

换句话说, 为计算 (简单) 随机变量的函数的数学期望, 没有必要直接利用全部概率测度 $\mathbf{P}$, 只需知道概率分布 P_ξ, 或等价地只需知道随机变量 ξ 的分布函数 F_ξ.

下面的定理归纳了这一性质.

定理 7 (勒贝格积分中的变量替换)　设 $(\Omega,\mathscr{F})$ 和 $(E,\mathscr{E})$ 是两个可测空间. $X = X(\omega)$ 是以 E 为值域的 $\mathscr{F}/\mathscr{E}-$ 可测函数. $\mathbf{P}$ 是 $(\Omega,\mathscr{F})$ 上的概率测度, P_X 是由 $X = X(\omega)$ 诱导的 $(E,\mathscr{E})$ 上的概率测度:

$$P_X(A) = \mathbf{P}\{\omega : X(\omega) \in A\}, A \in \mathscr{E}. \tag{40}$$

那么, 对于任意 $\mathscr{E}-$ 可测函数 $g = g(x), x \in E$,

$$\int_A g(x)P_X(\mathrm{d}x) = \int_{X^{-1}(A)} g(X(\omega))\mathbf{P}(\mathrm{d}\omega), \quad A \in \mathscr{E} \tag{41}$$

(其含义是: 如果有一个积分存在, 则另一个也存在, 并且二者相等).

证明　设集合 $A \in \mathscr{E}$ 和 $g(x) = I_B(x)$, 其中 $B \in \mathscr{E}$. 那么, 欲证的关系式 (41) 变为等式:

$$P_X(AB) = \mathbf{P}(X^{-1}(A) \cap X^{-1}(B)), \tag{42}$$

其正确性由 (40) 式和 $X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B)$ 可见.

由 (42) 式可见, (41) 式对于非负简单函数 $g = g(x)$ 成立, 而由单调收敛定理知, (41) 式对于任意非负 $\mathscr{E}-$ 可测函数成立.

而对于一般情形, 需要将函数 g 表示为 $g^+ - g^-$; 注意到, 因为 (41) 式对于函数 g^+ 和 g^- 成立, 而且例如, 若

$$\int_A g^+(x)P_X(\mathrm{d}x) < \infty,$$

则

$$\int_{X^{-1}(A)} g^+(X(\omega))\mathbf{P}(\mathrm{d}\omega) < \infty,$$

说明若

$$\int_A g(x)P_X(\mathrm{d}x)$$

存在, 则积分存在

$$\int_{X^{-1}(A)} g(X(\omega))\mathbf{P}(\mathrm{d}\omega). \qquad \square$$

系　$(E,\mathscr{E}) = (\mathbb{R},\mathscr{B}(\mathbb{R}))$, 而 $\xi = \xi(\omega)$ 是概率分布为 P_ξ 的随机变量. 那么, 如果 $g = g(x)$ 是博雷尔函数, 而积分

$$\int_A g(x)P_\xi(\mathrm{d}x) \quad 或 \quad \int_{\xi^{-1}(A)} g(\xi(\omega))\mathbf{P}(\mathrm{d}\omega)$$

存在, 则

$$\int_A g(x)P_\xi(\mathrm{d}x) = \int_{\xi^{-1}(A)} g(\xi(\omega))\mathbf{P}(\mathrm{d}\omega).$$

特别, 当 $A=\mathbb{R}$ 时得

$$\mathbf{E}g(\xi(\omega))=\int_{\Omega}g(\xi(\omega))\mathbf{P}(\mathrm{d}\omega)=\int_{\mathbb{R}}g(x)P_{\xi}(\mathrm{d}x). \tag{43}$$

测度 P_{ξ} 可以唯一地由分布函数还原 (§3 定理 1). 因此, 勒贝格积分

$$\int_{\mathbb{R}}g(x)P_{\xi}(\mathrm{d}x)$$

常记作

$$\int_{\mathbb{R}}g(x)F_{\xi}(\mathrm{d}x) \quad 或 \quad \int_{\mathbb{R}}g\mathrm{d}F_{\xi},$$

并称做 (关于分布函数 $F_{\xi}(x)$ 所对应的测度的) **勒贝格 - 斯蒂尔切斯积分**.

考虑分布函数 $F_{\xi}(x)$ 有密度 $f_{\xi}(x)$ 的情形, 即

$$F_{\xi}(x)=\int_{-\infty}^{x}f_{\xi}(y)\mathrm{d}y, \tag{44}$$

其中 $f_{\xi}=f_{\xi}(x)$ 非负博雷尔函数, 而积分是集合 $(-\infty,x]$ 上关于勒贝格测度的勒贝格积分 (见第 2 小节注 3). 在 (44) 式的记号下, (43) 式有如下形式:

$$\mathbf{E}g(\xi(\omega))=\int_{-\infty}^{\infty}g(x)f_{\xi}(x)\mathrm{d}x, \tag{45}$$

其中的积分是函数 $g(x)f_{\xi}(x)$ 按勒贝格测度的勒贝格积分. 事实上, 如果 $g(x)=I_B(x), B\in\mathscr{B}(\mathbb{R})$, 则 (45) 式有如下形式:

$$P_{\xi}(B)=\int_{B}f_{\xi}(x)\mathrm{d}x, B\in\mathscr{B}(\mathbb{R}), \tag{46}$$

其正确性由 §3 的定理 1 和如下公式可见:

$$F_{\xi}(b)-F_{\xi}(a)=\int_{a}^{b}f_{\xi}(x)\mathrm{d}x.$$

一般情形的证明, 与定理 7 的证明相同.

10. 傅比尼 (I. L. Fubini) **定理** 考虑带测度 μ 的可测空间 $(\Omega,\mathscr{F})$, 其中 $\Omega=\Omega_1\times\Omega_2,\mathscr{F}=\mathscr{F}_1\otimes\mathscr{F}_2$, 而测度 $\mu=\mu_1\times\mu_2$ 是有限测度 μ_1 和 μ_2 的直积, 即在 $\mathscr{F}$ 上的测度:

$$\mu_1\times\mu_2(A\times B)=\mu_1(A)\mu_2(B), \quad A\in\mathscr{F}_1, B\in\mathscr{F}_2$$

(由下面将要证明的定理 8, 可见这样测度的存在性).

下面的定理有重要意义, 就像数学分析中关于 "**二重黎曼积分化为二次积分**" 的著名定理一样.

定理 8 (傅比尼定理)　设 $\xi=\xi(\omega_1,\omega_2)$ 是 $\mathscr{F}_1\otimes\mathscr{F}_2-$ 可测函数, 并且关于测度 $\mu_1\times\mu_2$ 可积:

$$\int_{\Omega_1\times\Omega_2}|\xi(\omega_1,\omega_2)|\mathrm{d}(\mu_1\times\mu_2)<\infty. \tag{47}$$

那么, 积分

$$\int_{\Omega_1}\xi(\omega_1,\omega_2)\mu_1(\mathrm{d}\omega_1)\quad 和\quad \int_{\Omega_2}\xi(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)$$

1) 对于 μ_2- 几乎一切 ω_2 和 μ_1- 几乎一切 ω_1 有定义;

2) 是 $\mathscr{F}_1-$ 和 $\mathscr{F}_2-$ 可测函数, 并且相应地

$$\begin{aligned}\mu_2\left\{\omega_2:\int_{\Omega_1}|\xi(\omega_1,\omega_2)|\mu_1(\mathrm{d}\omega_1)=\infty\right\}=0,\\ \mu_1\left\{\omega_1:\int_{\Omega_2}|\xi(\omega_1,\omega_2)|\mu_2(\mathrm{d}\omega_2)=\infty\right\}=0;\end{aligned} \tag{48}$$

3)

$$\begin{aligned}\int_{\Omega_1\times\Omega_2}\xi(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2)&=\int_{\Omega_1}\left[\int_{\Omega_2}\xi(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1)\\ &=\int_{\Omega_2}\left[\int_{\Omega_1}\xi(\omega_1,\omega_2)\mu_1(\mathrm{d}\omega_1)\right]\mu_2(\mathrm{d}\omega_2).\end{aligned} \tag{49}$$

证明　(i) 首先, 证明对于任意固定 $\omega_1\in\Omega_1$,ω_2 的函数 $\xi_{\omega_1}(\omega_2)=\xi(\omega_1,\omega_2)$ 是 $\mathscr{F}_2-$ 可测的.

设 $F\in\mathscr{F}_1\otimes\mathscr{F}_2$ 和 $\xi(\omega_1,\omega_2)=I_F(\omega_1,\omega_2)$. 记 $F_{\omega_1}=\{\omega_2\in\Omega_2:(\omega_1,\omega_2)\in F\}$ 为 F 在点 ω_1 处的截线, 并设 $\mathscr{C}_{\omega_1}=\{F\in\mathscr{F}:F_{\omega_1}\in\mathscr{F}_2\}$. 需要证明对于任意 ω_1,$\mathscr{C}_{\omega_1}\in\mathscr{F}$.

如果 $F=A\times B,A\in\mathscr{F}_1,B\in\mathscr{F}_2$, 则

$$(A\times B)_{\omega_1}=\begin{cases}B,\ \omega_1\in A,\\ \varnothing,\ \omega_1\notin A.\end{cases}$$

因此, 具有可测边的矩形属于 $\mathscr{C}_{\omega_1}$. 其次, 如果 $F\in\mathscr{F}$, 则 $(\overline{F})_{\omega_1}=\overline{F_{\omega_1}}$, 而若 $\{F^n\}_{n\geqslant1}$ 是 F 中的集合, 则 $(\bigcup F^n)_{\omega_1}=\bigcup F^n_{\omega_1}$. 由此可见 $\mathscr{C}_{\omega_1}\in\mathscr{F}$.

现在假设 $\xi(\omega_1,\omega_2)\geqslant0$. 那么, 由于对于每一 ω_1, 函数 $\xi_{\omega_1}(\omega_2)=\xi(\omega_1,\omega_2)$ 是 $\mathscr{F}_2-$ 可测的, 则积分

$$\int_{\Omega_2}\xi(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)$$

有定义. 现在证明该积分是 $\mathscr{F}_1-$ 可测函数, 且

$$\int_{\Omega_1}\left[\int_{\Omega_2}\xi(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1)=\int_{\Omega_1\times\Omega_2}\xi(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2). \tag{50}$$

假设 $\xi(\omega_1,\omega_2) = I_{A\times B}(\omega_1,\omega_2), A \in \mathscr{F}_1, B \in \mathscr{F}_2$, 则由于 $I_{A\times B}(\omega_1,\omega_2) = I_A(\omega_1)I_B(\omega_2)$, 可见

$$\int_{\Omega_2} I_{A\times B}(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2) = I_A(\omega_1)\int_{\Omega_2} I_B(\omega_2)\mu_2(\mathrm{d}\omega_2). \tag{51}$$

因而, (51) 式左侧的积分是 $\mathscr{F}_1-$ 可测函数.

设 $\xi(\omega_1,\omega_2) = I_F(\omega_1,\omega_2), F \in \mathscr{F} = \mathscr{F}_1 \otimes \mathscr{F}_2$, 证明积分

$$f(\omega_1) = \int_{\Omega_2} I_F(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)$$

$\mathscr{F}-$ 可测. 为此, 记 $\mathscr{C} = \{F \in \mathscr{F} : f(\omega_1)$ 为 $\mathscr{F}_1-$ 可测$\}$. 由于已经证明集合 $A\times B$ 属于 $\mathscr{C}$, $(A \in \mathscr{F}_1, B \in \mathscr{F}_2)$, 说明由这样集合组成的代数 $\mathscr{A}$ 也属于 $\mathscr{C}$. 由单调收敛定理可见, 集系 $\mathscr{C}$ 是单调类, 故 $\mathscr{C} = \mu(\mathscr{C})$ 是含 $\mathscr{C}$ 的最小单调类. 所以, 由于包含关系 $\mathscr{A} \subseteq \mathscr{C} \subseteq \mathscr{F}$, 以及 §2 定理 1, 可见 $\mathscr{F} = \sigma(\mathscr{A}) = \mu(\mathscr{A}) \subseteq \mu(\mathscr{C}) = \mathscr{C} \subseteq \mathscr{F}$, 即 $\mathscr{C} = \mathscr{F}$.

最后, 如果 $\xi(\omega_1,\omega_2)$ 是任意非负 $\mathscr{F}-$ 可测函数, 则由单调收敛定理和 §4 定理 2, 可见下面积分的 $\mathscr{F}_1-$ 可测性:

$$\int_{\Omega_2} \xi(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2).$$

现在证明, 定义在 $\mathscr{F} = \mathscr{F}_1 \otimes \mathscr{F}_2$ 上, 且具有性质: $\mu = \mu_1 \times \mu_2(A\times B) = \mu_1(A)\mu_2(B), A \in \mathscr{F}_1, B \in \mathscr{F}_2$, 的测度 $\mu = \mu_1 \times \mu_2$, 测度确实存在而且唯一. 对于 $F \in \mathscr{F}$,

$$\mu(F) = \int_{\Omega_1}\left[\int_{\Omega_2} I_{F_{\omega_1}}(\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1).$$

如已经证明的, 上面的二次积分中里边的积分是 $\mathscr{F}_1-$ 可测函数, 因而集函数 $\mu(F)$ 确实对 $F \in \mathscr{F}$ 有定义. 显然, 若 $F = A\times B$, 则 $\mu(A\times B) = \mu_1(A)\mu_2(B)$. 设 $\{F^n\}$ 是 $\mathscr{F}$ 中两两不相交的集合, 则

$$\begin{aligned}
\mu\left(\sum_n F^n\right) &= \int_{\Omega_1}\left[\int_{\Omega_2} I_{(\sum F^n)_{\omega_1}}(\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1)\\
&= \int_{\Omega_1}\sum_n\left[\int_{\Omega_2} I_{F^n_{\omega_1}}(\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1)\\
&= \sum_n\int_{\Omega_1}\left[\int_{\Omega_2} I_{F^n_{\omega_1}}(\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1) = \sum_n \mu(F^n).
\end{aligned}$$

于是, μ 是 $\mathscr{F}$ 上的 ($\sigma-$ 有限) 测度.

由卡拉泰奥多里定理, 可见该测度是唯一具有性质 $\mu(A\times B) = \mu_1(A)\mu_2(B)$ 的测度.

现在证明 (50) 式. 如果 $\xi(\omega_1,\omega_2)=I_{A\times B}(\omega_1,\omega_2), A\in\mathscr{F}_1, B\in\mathscr{F}_2$, 则

$$\int_{\Omega_1\times\Omega_2} I_{A\times B}(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2)=\mu_1\times\mu_2(A\times B), \tag{52}$$

而由于 $I_{A\times B}(\omega_1,\omega_2)=I_A(\omega_1)I_B(\omega_2)$, 可见

$$\begin{aligned}&\int_{\Omega_1}\left[\int_{\Omega_2} I_{A\times B}(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1)\\ =&\int_{\Omega_1}\left[I_A(\omega_1)\int_{\Omega_2} I_B(\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1)=\mu_1(A)\mu_2(B).\end{aligned} \tag{53}$$

根据测度 $\mu_1\times\mu_2$ 的定义

$$\mu_1\times\mu_2(A\times B)=\mu_1(A)\mu_2(B),$$

由 (52) 和 (53) 式, 可见对于 $\xi(\omega_1,\omega_2)=I_{A\times B}(\omega_1,\omega_2)$, 式 (50) 成立.

现在设 $\xi(\omega_1,\omega_2)=I_F(\omega_1,\omega_2), F\in\mathscr{F}$. 集合的函数

$$\lambda(F)=\int_{\Omega_1\times\Omega_2} I_F(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2), F\in\mathscr{F},$$

显然是 $\sigma-$ 有限测度. 不难验证, 集函数

$$\nu(F)=\int_{\Omega_1}\left[\int_{\Omega_2} I_F(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1)$$

也是这样的测度. 上面已经证明, 测度 λ 和 ν 在形如 $F=A\times B$ 的集合上重合, 因此也在代数 $\mathscr{A}$ 上重合. 由此根据卡拉泰奥多里定理, 可见测度 λ 和 ν 对于一切 $F\in\mathscr{F}$ 重合.

(ii) 现在证明傅比尼定理本身的命题. 由 (47) 式, 有

$$\int_{\Omega_1\times\Omega_2}\xi^+(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2)<\infty, \int_{\Omega_1\times\Omega_2}\xi^-(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2)<\infty.$$

由已证明的知, 积分

$$\int_{\Omega_2}\xi^+(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)$$

是 ω_1 的 $\mathscr{F}_1-$ 可测函数, 且

$$\int_{\Omega_1}\left[\int_{\Omega_2}\xi^+(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)\right]\mu_1(\mathrm{d}\omega_1)=\int_{\Omega_1\times\Omega_2}\xi^+(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2)<\infty.$$

因此, 由练习题 4 (亦见第 3 小节的性质 **J**)

$$\int_{\Omega_2}\xi^+(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)<\infty(\mu_1-\text{a.c.}).$$

同样,

$$\int_{\Omega_2} \xi^-(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2) < \infty(\mu_1 - \text{a.c.}),$$

从而,

$$\int_{\Omega_2} |\xi(\omega_1,\omega_2)|\mu_2(\mathrm{d}\omega_2) < \infty(\mu_1 - \text{a.c.}).$$

显然, 除测度 μ_1 为 0 的集合 $\mathscr{N}$ 外, 有

$$\int_{\Omega_2} \xi(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2) = \int_{\Omega_2} \xi^+(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2) - \int_{\Omega_2} \xi^-(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2). \tag{54}$$

假设其中对于 $\omega_1 \in \mathscr{N}$, 积分等于 0, 则可以认为 (54) 式对于一切 $\omega_1 \in \Omega_1$ 成立. 那么, 考虑到 (50) 式, 将 (54) 式对测度 μ_1 积分, 得

$$\begin{aligned}
&\int_{\Omega_1} \left[\int_{\Omega_2} \xi(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)\right] \mu_1(\mathrm{d}\omega_1) \\
&= \int_{\Omega_1} \left[\int_{\Omega_2} \xi^+(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)\right] \mu_1(\mathrm{d}\omega_1) - \int_{\Omega_1} \left[\int_{\Omega_2} \xi^-(\omega_1,\omega_2)\mu_2(\mathrm{d}\omega_2)\right] \mu_1(\mathrm{d}\omega_1) \\
&= \int_{\Omega_1\times\Omega_2} \xi^+(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2) - \int_{\Omega_1\times\Omega_2} \xi^-(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2) \\
&= \int_{\Omega_1\times\Omega_2} \xi(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2).
\end{aligned}$$

类似地证明 (48) 式第一个关系式和下面的等式:

$$\int_{\Omega_1\times\Omega_2} \xi(\omega_1,\omega_2)\mathrm{d}(\mu_1\times\mu_2) = \int_{\Omega_2} \left[\int_{\Omega_1} \xi(\omega_1,\omega_2)\mu_1(\mathrm{d}\omega_1)\right] \mu_2(\mathrm{d}\omega_2). \qquad \square$$

系 *如果*

$$\int_{\Omega_1} \left[\int_{\Omega_2} |\xi(\omega_1,\omega_2)|\mu_2(\mathrm{d}\omega_2)\right] \mu_1(\mathrm{d}\omega_1) < \infty,$$

则傅比尼定理的结论仍然成立.

事实上, 由条件 (50) 可见 (47) 式, 因此傅比尼定理的一切结论仍然成立.

例 设随机向量 (ξ,η) 有二维概率密度 $f_{\xi,\eta}(x,y)$, 即

$$\mathbf{P}\{(\xi,\eta)\in B\} = \int_B f_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y, B\in\mathscr{B}(\mathbb{R}^2),$$

其中 $f_{\xi,\eta}(x,y)$ 是非负 $\mathscr{B}(\mathbb{R}^2)$– 可测函数, 而积分是对二维勒贝格测度的积分.

我们证明, ξ 和 η 的一维概率分布也有密度 $f_\xi(x)$ 和 $f_\eta(y)$, 并且

$$f_\xi(x) = \int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\mathrm{d}y, f_\eta(y) = \int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\mathrm{d}x, \tag{55}$$

事实上, 如果 $A\in\mathscr{B}(\mathbb{R})$, 则根据傅比尼定理

$$\mathbf{P}\{\xi\in A\} = \mathbf{P}\{(\xi,\eta)\in A\times\mathbb{R}\} = \int_{A\times\mathbb{R}} f_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y = \int_A \left[\int_{\mathbb{R}} f_{\xi,\eta}(x,y)\mathrm{d}y\right] \mathrm{d}x,$$

因此, 说明 ξ 有概率分布密度, 而 (55) 式的第一个公式得证. 类似地可以证明 (55) 式的第二个公式.

根据 §5 的定理, 随机变量 ξ 和 η 独立的充分和必要条件是

$$F_{\xi,\eta}(x,y)=F_{\xi}(x)F_{\eta}(y),(x,y)\in\mathbb{R}^2.$$

现在证明, 当随机变量 ξ 和 η 有联合密度 $f_{\xi,\eta}(x,y)$ 时, 它们独立的充分必要条件是:

$$f_{\xi,\eta}(x,y)=f_{\xi}(x)f_{\eta}(y) \tag{56}$$

(等式应理解为关于二维勒贝格测度几乎必然成立).

实际上, 如果 (56) 式成立, 则根据傅比尼定理, 有

$$\begin{aligned}F_{\xi,\eta}(x,y)&=\int_{(-\infty,x]\times(-\infty,y]}f_{\xi,\eta}(s,t)\mathrm{d}s\mathrm{d}t=\int_{(-\infty,x]\times(-\infty,y]}f_{\xi}(s)f_{\eta}(t)\mathrm{d}s\mathrm{d}t\\&=\int_{(-\infty,x]}f_{\xi}(s)\mathrm{d}s\left[\int_{(-\infty,y]}f_{\eta}(t)\mathrm{d}t\right]=F_{\xi}(x)F_{\eta}(y),\end{aligned}$$

从而 ξ 和 η 独立.

相反, 如果 ξ 和 η 独立, 且有联合密度 $f_{\xi,\eta}(x,y)$, 则仍根据傅比尼定理, 有

$$\begin{aligned}\int_{(-\infty,x]\times(-\infty,y]}f_{\xi,\eta}(s,t)\mathrm{d}s\mathrm{d}t&=\left[\int_{(-\infty,x]}f_{\xi}(s)\mathrm{d}s\right]\times\left[\int_{(-\infty,y]}f_{\eta}(t)\mathrm{d}t\right]\\&=\int_{(-\infty,x]\times(-\infty,y]}f_{\xi}(s)f_{\eta}(t)\mathrm{d}s\mathrm{d}t.\end{aligned}$$

由此可见, 对于任意 $B\in\mathscr{B}(\mathbb{R}^2)$, 有

$$\int_B f_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y=\int_B f_{\xi}(x)f_{\eta}(y)\mathrm{d}x\mathrm{d}y,$$

则有性质 **I**, 易见条件 (56) 成立.

11. 勒贝格积分和黎曼积分的各种定义及其关系 这一小节, 将要讨论勒贝格积分和黎曼积分各种定义的问题, 以及它们之间的关系.

首先需要指出, 勒贝格积分的构造, 与需要积分的函数定义在什么可测空间 $(\Omega,\mathscr{F})$ 无关. 然而黎曼积分, 对于抽象函数根本没有定义, 对于空间 $\Omega=\mathbb{R}^n$ 的情形, 则依次定义: 首先对 $\mathbb{R}^1$ 定义, 然后随着相应的变化, 依次对于 $n>1$ 的情形进行移植.

应该强调, 黎曼积分和勒贝格积分的构造, 基于不同的思想基础. 构造黎曼积分的第一步, 将 $x\in\mathbb{R}^1$ 点按其在 x 轴上临近程度的特征进行分组. 而在构造勒贝格积分时, 则将 $x\in\mathbb{R}^1$ 点按另一种特征分组: 按被积函数值的临近程度进行分组. 这两种不同处理方法的结果, 使仅对于不 "十分" 间断的函数, 黎曼积分的积分和才有极限; 然而, 勒贝格积分的积分和, 却对更广泛的函数类收敛于极限值.

回忆黎曼 - 斯蒂尔切斯积分的定义. 假设 $G(x)$ 是 $\mathbb{R}^1$ 上的某一广义分布函数 (见 §3 第 2 小节), μ 是与其对应的勒贝格 - 斯蒂尔切斯测度, 而 $g = g(x)$ 是在区间 $[a,b]$ 之外为 0 的有界函数.

考虑区间 $[a,b]$ 的分割 $\mathscr{P} = \{x_0, \cdots, x_n\}$, 其中

$$a = x_0 < x_1 < \cdots < x_n = b,$$

建立上积分和与下积分和:

$$\overline{\sum_{\mathscr{P}}} = \sum_{i=1}^{n} \overline{g}_i[G(x_{i+1}) - G(x_i)], \quad \underline{\sum_{\mathscr{P}}} = \sum_{i=1}^{n} \underline{g}_i[G(x_{i+1}) - G(x_i)],$$

其中

$$\overline{g}_i = \sup_{x_{i-1}<y\leqslant x_i} g(y), \quad \underline{g}_i = \inf_{x_{i-1}<y\leqslant x_i} g(y).$$

定义简单函数 $\overline{g}_{\mathscr{P}}(x)$ 和 $\underline{g}_{\mathscr{P}}(x)$, 并对于 $x_{i-1} < x \leqslant x_i$, 设

$$\overline{g}_{\mathscr{P}}(x) = \overline{g}_i, \underline{g}_{\mathscr{P}}(x) = \underline{g}_i,$$

并设 $\overline{g}_{\mathscr{P}}(a) = \underline{g}_{\mathscr{P}}(a) = g(a)$. 那么显然, 与根据勒贝格 - 斯蒂尔切斯积分的构造 (见第 2 小节的注 3), 有

$$\overline{\sum_{\mathscr{P}}} = (\mathrm{L-S}) \int_a^b \overline{g}_{\mathscr{P}}(x) G(\mathrm{d}x)$$

和

$$\underline{\sum_{\mathscr{P}}} = (\mathrm{L-S}) \int_a^b \overline{g}_{\mathscr{P}}(x) G(\mathrm{d}x).$$

现在设 $\{\mathscr{P}_k\}$ 是分割序列, 满足 $\mathscr{P}_k \subseteq \mathscr{P}_{k+1}$, 且对于 $\mathscr{P}_k = \{x_0^{(k)}, \cdots, x_{n_k}^{(k)}\}$, 当 $k \to \infty$ 时 $\max\limits_{0\leqslant i\leqslant n_k} |x_{i+1}^{(k)} - x_i^{(k)}| \to 0$. 那么, $\overline{g}_{\mathscr{P}_1} \geqslant \overline{g}_{\mathscr{P}_2} \geqslant \cdots \geqslant g \geqslant \cdots \geqslant \underline{g}_{\mathscr{P}_2} \geqslant \underline{g}_{\mathscr{P}_1}$, 且如果 $|g(x)| \leqslant C$, 则根据控制收敛定理, 有

$$\begin{aligned} \lim_{k\to\infty} \overline{\sum_{\mathscr{P}_k}} &= (\mathrm{L-S}) \int_a^b \overline{g}(x) G(\mathrm{d}x), \\ \lim_{k\to\infty} \underline{\sum_{\mathscr{P}_k}} &= (\mathrm{L-S}) \int_a^b \underline{g}(x) G(\mathrm{d}x), \end{aligned} \tag{57}$$

其中 $\overline{g}(x) = \lim_k \overline{g}_{\mathscr{P}_k}(x), \underline{g}(x) = \lim_k \underline{g}_{\mathscr{P}_k}(x)$.

如果极限 $\lim_k \overline{\sum}_{\mathscr{P}_k}$ 和 $\lim_k \underline{\sum}_{\mathscr{P}_k}$ 有限, 相等, 其共同值与分割序列 $\{\mathscr{P}_k\}$ 的选择无关, 则称函数 $g = g(x)$ 黎曼-斯蒂尔切斯可积, 而它们极限的共同值记作

$$(\mathrm{R-S}) \int_a^b g(x) G(\mathrm{d}x), \quad 或 \quad (\mathrm{R-S}) \int_a^b g(x) \mathrm{d}G(x). \tag{58}$$

当 $G(x) = x$ 时, 该积分称做黎曼积分, 记作

$$(\mathrm{R}) \int_a^b g(x) \mathrm{d}x.$$

现在假设

$$(\mathrm{L-S}) \int_a^b g(x) G(\mathrm{d}x)$$

是相应的勒贝格 - 斯蒂尔切斯积分 (见第 2 小节, 注 3).

定理 9 如果函数 $g = g(x)$ 在 $[a,b]$ 上连续, 则它黎曼 - 斯蒂尔切斯可积, 且

$$(\mathrm{R-S}) \int_a^b g(x) G(\mathrm{d}x) = (\mathrm{L-S}) \int_a^b g(x) G(\mathrm{d}x). \tag{59}$$

证明 由于函数 $g = g(x)$ 连续, 可见 $\overline{g}(x) = g(x) = \underline{g}(x)$. 因此, 由 (57) 式, 有

$$\lim_{k\to\infty} \overline{\sum_{\mathscr{P}_k}} = \lim_{k\to\infty} \underline{\sum_{\mathscr{P}_k}}.$$

于是, 连续函数 $g = g(x)$ 黎曼 - 斯蒂尔切斯可积, 并且 (仍然由于 (57) 式) 其积分等于勒贝格 - 斯蒂尔切斯积分. □

对于直线 $\mathbb{R}$ 上勒贝格测度的情形, 我们比较详细地讨论黎曼积分和勒贝格积分之间的关系问题.

定理 10 假设 $g = g(x)$ 在 $[a,b]$ 上是有界函数.

a) 函数 $g = g(x)$ 在 $[a,b]$ 上黎曼可积, 当且仅当它 (关于 $\overline{\mathscr{B}}([a,b])$ 上的勒贝格 $\overline{\lambda}$ 测度) 几乎处处连续.

b) 假如函数 $g = g(x)$ 在 $[a,b]$ 上黎曼可积, 则它也勒贝格可积, 且

$$(\mathrm{R}) \int_a^b g(x) \mathrm{d}x = (\mathrm{L}) \int_a^b g(x) \overline{\lambda}(\mathrm{d}x). \tag{60}$$

证明 a) 假设函数 $g = g(x)$ 黎曼可积. 那么, 由 (57) 式, 可见

$$(\mathrm{L}) \int_a^b \overline{g}(x) \overline{\lambda}(\mathrm{d}x) = (\mathrm{L}) \int_a^b \underline{g}(x) \overline{\lambda}(\mathrm{d}x).$$

因为一般 $\underline{g}(x) \leqslant g(x) \leqslant \overline{g}(x)$, 所以由性质 **H**, 有

$$\underline{g}(x) = g(x) = \overline{g}(x) \quad (\overline{\lambda}-\text{a.c.}). \tag{61}$$

由此不难验证, 函数 $g = g(x)$ (关于测度 $\overline{\lambda}$) 几乎处处连续.

相反, 假设函数 $g = g(x)$ (关于测度 $\overline{\lambda}$) 几乎处处连续. 这时 (61) 式成立, 因此 $g(x)$ 仅在 $\overline{\lambda}(\mathscr{N}) = 0$ 的集合 $\mathscr{N}$ 上不同于函数 $\overline{g}(x)$. 那么

$$\begin{aligned}\{x : g(x) \leqslant c\} &= \{x : g(x) \leqslant c\} \cap \overline{\mathscr{N}} + \{x : g(x) \leqslant c\} \cap \mathscr{N} \\ &= \{x : \overline{g}(x) \leqslant c\} \cap \overline{\mathscr{N}} + \{x : g(x) \leqslant c\} \cap \mathscr{N}.\end{aligned}$$

显然, $\{x:\overline{g}(x)\leqslant c\}\cap\overline{\mathscr{N}}\in\overline{\mathscr{B}}([a,b])$, 而由于 $\{x:g(x)\leqslant c\}\cap\mathscr{N}$ 是 $\mathscr{N}$ 的子集, 而 $\mathscr{N}$ 是关于勒贝格测度 $\overline{\lambda}$ 的 0 测集, 故 $\mathscr{N}\in\overline{\mathscr{B}}([a,b])$. 从而, 函数 $g=g(x)$ 为 $\overline{\mathscr{B}}([a,b])-$ 可测的, 并且作为有界函数勒贝格可积. 所以根据性质 **G**, 有

$$(\mathrm{L})\int_a^b\overline{g}(x)\overline{\lambda}(\mathrm{d}x)=(\mathrm{L})\int_a^b\underline{g}(x)\overline{\lambda}(\mathrm{d}x)=(\mathrm{L})\int_a^b g(x)\overline{\lambda}(\mathrm{d}x).$$

于是, 命题 a) 得证.

b) 如果函数 $g=g(x)$ 黎曼可积, 则由 a) 知它连续 ($\overline{\lambda}$–a.c.). 由定理 9 知, 对于连续函数, $g=g(x)$ 也勒贝格可积, 并且黎曼积分与勒贝格积分相等. □

注 1 设 μ 是 $\mathscr{B}([a,b])$ 上的某一勒贝格 - 斯蒂尔切斯测度. 对于子集 $\Lambda\subseteq[a,b]$ 的集系 $\overline{\mathscr{B}}_\mu([a,b])$, 存在集合 $A,B\in\mathscr{B}([a,b])$, 使 $A\subseteq\Lambda\subseteq B,\mu(B\backslash A)=0$. 设 $\overline{\mu}$ 是**测度 μ 到 $\mathscr{B}([a,b])$ 上的开拓** (对于使 $A\subseteq\Lambda\subseteq B,\mu(B\backslash A)=0$ 的 $\Lambda,\overline{\mu}(\Lambda)=\mu(A)$). 那么, 如果用测度 $\overline{\mu}$ 代替勒贝格测度 $\overline{\lambda}$, 而相应地用黎曼 – 斯蒂尔切斯积分和勒贝格 - 斯蒂尔切斯积分, 分别代替黎曼积分和勒贝格积分, 则定理仍然成立.

注 2 勒贝格积分定义 (见第 1 小节定义 1 和 2, 以及公式 (3) 和 (6)) 既从概念上, 又 "纯粹表面地" 区别于黎曼和黎曼 - 斯蒂尔切斯积分, 而后两种积分用到**上、下积分和** (见 (57) 式).

现在较详细地比较这些定义.

设 $(\Omega,\mathscr{F},\mu)$ 是某一带测度 μ 的可测空间.

对于每一 $\mathscr{F}-$ 可测非负函数 $f=f(\omega)$, 定义两个 (下和上) 积分 L_*f 和 L^*f (记作 $\int_* f\mathrm{d}\mu$ 和 $\int^* f\mathrm{d}\mu$):

$$L_*f=\sup\sum_i\left(\inf_{\omega\in A_i}f(\omega)\right)\mu(A_i),$$

$$L^*f=\inf\sum_i\left(\sup_{\omega\in A_i}f(\omega)\right)\mu(A_i),$$

其中 sup 和 inf 对空间 Ω 的一切 $\mathscr{F}-$ 可测有限分割 $(A_1,A_2,\cdots,A_n),A_1,A_2,\cdots,A_n$ 是 $\mathscr{F}-$ 可测集合, $(A_1+A_2+\cdots+A_n)=\Omega,n\geqslant1$.

可以证明 $L_*f\leqslant L^*f$, 而如果函数 f 有界且测度 μ 有限, 则 $L_*f=L^*f$ (练习题 20).

定义函数 f 对测度 μ 的积分 Lf 的方法之一 (达布 – 扬 [J. G. Darboux-W. H. Young]), **称函数 f 对测度 μ 是可积的, 如果 $L_*f=L^*f$, 并在这种情况下记作** $Lf=L_*f(=L^*f)$.

如果现在转向在第 1 小节给出的勒贝格积分 $\mathbf{E}f$ 的定义 1, 则可以看到 (练习题 21)

$$\mathbf{E}f=L_*f.$$

于是, 可以认为, 对于有界非负函数 $f = f(\omega)$, 勒贝格方法和达布 – 扬方法导致同一结果 ($\mathbf{E}f = Lf = L_*f = L^*f$).

这些积分方法的差异出现在如下情形中: 被积函数无界, 或者测度μ 无限.

例如, 按勒贝格积分的观点, 相应地对于 $f(x) = x^{-1/2}I_{(0,1]}$ 和 $f(x) = x^{-2}I_{(1,\infty)}$, 积分

$$\int_{(0,1]} \frac{\mathrm{d}x}{x^{1/2}} \quad 和 \quad \int_{(1,\infty)} \frac{\mathrm{d}x}{x^2}$$

相等, 且都等于 L_*f. 然而, $L^*f = \infty$.

这样, $L_*f < L^*f$, 说明所考虑的函数在达布 – 扬意义上不可积, 但在勒贝格意义上可积.

在所阐述的、运用下积分 L_*f 及上积分 L^*f 方法的范围内, 我们转向黎曼意义上的积分法.

假设 $\Omega = (0,1], \mathscr{F} = \mathscr{B}$, (博雷尔 $\sigma-$ 代数), 而 $\mu = \lambda$ 是勒贝格测度. 设 $f = f(x), x \in \Omega$ 是某一有界函数 (关于其可测性暂时不作假设).

仿照 L_*f 和 L^*f, 引进下黎曼积分 R_*f 和上黎曼积分 R^*f, 设

$$R_*f = \sup \sum_i \left(\inf_{\omega \in B_i} f(\omega) \right) \lambda(B_i),$$

$$R^*f = \inf \sum_i \left(\sup_{\omega \in B_i} f(\omega) \right) \lambda(B_i),$$

其中 $(B_1, B_2, \cdots, B_n)$ 是 $\Omega = (0,1]$ 的有限分割, 而 B_i 是形如 $(a_i, b_i]$ 的集合 (B_i 与 L_*f 和 L^*f 定义中 A_i 的不同, A_i 是任意 $\mathscr{F}-$ 可测集合).

由上面引进的定义, 显然

$$R_*f \leqslant L_*f \leqslant L^*f \leqslant R^*f.$$

在定理 9 和 10 中引进的黎曼可积性的性质, 可以重新表述, 并充实下列条件:

(a) $R^*f = R_*F$;

(b) 函数 f 之间断点集合 D_f 的勒贝格测度等于 $0: \lambda(D_f) = 0$;

(c) 存在一常数 $R(f)$, 使对于任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使得

$$\left| R(f) - \sum_i f(\omega_i) \lambda((a_i, b_i]) \right| < \varepsilon$$

且对于两两不相交区间 $(a_i, b_i], \sum (a_i, b_i] = (0,1]$ 的任何有限族, 有 $\lambda((a_i, b_i]) < \delta, \omega_i \in (a_i, b_i]$.

利用定理 9 和 10 的论证, 可以证明 (练习题 22), 如果函数 f 有界, 则

(A) 条件 (a),(b),(c) 等价, 并且

(B) 在 (a),(b),(c) 中的任何条件下, 有

$$R(f) = R_*f = R^*f.$$

12. 勒贝格 – 斯蒂尔切斯积分的分部积分法 在这一小节, 我们引进关于"勒贝格 – 斯蒂尔切斯积分中分部积分"的有用的定理.

设在 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上有两个广义分布函数 $F=F(x)$ 和 $G=G(x)$.

定理 11 对于任意实数 a 和 $b, a<b$, 如下分部积分公式成立:

$$F(b)G(b)-F(a)G(a)=\int_a^b F(s-)\mathrm{d}G(s)+\int_a^b G(s)\mathrm{d}F(s), \tag{62}$$

或等价地,

$$\begin{aligned}&F(b)G(b)-F(a)G(a)\\&=\int_a^b F(s-)\mathrm{d}G(s)+\int_a^b G(s-)\mathrm{d}F(s)+\sum_{a<s\leqslant b}\Delta F(s)\Delta G(s),\end{aligned} \tag{63}$$

其中

$$F(s-)=\lim_{t\uparrow s}F(t), \Delta F(s)=F(s)-F(s-).$$

注 1 符号公式 (62) 可以写成如下"微分"形式:

$$\mathrm{d}(FG)=F_-\mathrm{d}G+G\mathrm{d}F. \tag{64}$$

注 2 对于 ($[a,b]$ 上的) 有界变差函数 F 和 G, 定理的结论仍然成立. (每一个这样的右连续且有左极限的函数, 可以表示为两个单调不减函数之差.)

证明 首先回忆, 根据第 2 小节的假设, 将 $\int_a^b$ 积分理解为 $\int_{(a,b]}$. 因此 (见 §3 中公式 (2)):

$$[F(b)-F(a)][G(b)-G(a)]=\int_a^b \mathrm{d}F(s)\times\int_a^b \mathrm{d}G(t).$$

由此根据傅比尼定理 ($F\times G$ 是对应于 F 和 G 的测度的直积), 可见

$$\begin{aligned}&[F(b)-F(a)][G(b)-G(a)]=\int_{(a,b]\times(a,b]}\mathrm{d}(F\times G)(s,t)\\&=\int_{(a,b]\times(a,b]}I_{\{s\geqslant t\}}(s,t)\mathrm{d}(F\times G)(s,t)+\int_{(a,b]\times(a,b]}I_{\{s<t\}}(s,t)\mathrm{d}(F\times G)(s,t)\\&=\int_{(a,b]}[G(s)-G(a)]\mathrm{d}F(s)+\int_{(a,b]}[F(t-)-F(a)]\mathrm{d}G(t)\\&=\int_a^b G(s)\mathrm{d}F(s)+\int_a^b F(s-)\mathrm{d}G(s)-G(a)[F(b)-F(a)]\\&-F(a)[G(b)-G(a)],\end{aligned} \tag{65}$$

其中 I_A 是集合 A 的示性函数.

由 (65) 式直接可得 (62) 式, 同样, 由 (62) 式得 (63) 式, 为此只需注意到

$$\int_a^b [G(s)-G(s-)]\mathrm{d}F(s)=\sum_{a<s\leqslant b}\Delta G(s)\Delta F(s). \tag{66}$$

□

系 1 如果 $F(x)$ 和 $G(x)$ 是分布函数, 则

$$F(x)G(x)=\int_{-\infty}^{x}F(s-)\mathrm{d}G(s)+\int_{-\infty}^{x}G(s)\mathrm{d}F(s). \tag{67}$$

如果分布函数

$$F(x)=\int_{-\infty}^{x}f(s)\mathrm{d}s,$$

则

$$F(x)G(x)=\int_{-\infty}^{x}F(s)\mathrm{d}G(s)+\int_{-\infty}^{x}G(s)f(s)\mathrm{d}s. \tag{68}$$

系 2 设随机变量 ξ 的分布函数为 $F=F(x)$, 而且 $\mathbf{E}|\xi|^n<\infty$. 那么,

$$\int_0^{\infty}x^n\mathrm{d}F(x)=n\int_0^{\infty}x^{n-1}[1-F(x)]\mathrm{d}x, \tag{69}$$

$$\int_{-\infty}^{0}|x|^n\mathrm{d}F(x)=n\int_0^{\infty}x^{n-1}F(-x)\mathrm{d}x, \tag{70}$$

$$\mathbf{E}|\xi|^n=\int_{-\infty}^{\infty}|x|^n\mathrm{d}F(x)=n\int_0^{\infty}x^{n-1}[1-F(x)+F(-x)]\mathrm{d}x. \tag{71}$$

为证明 (69) 式, 注意到

$$\begin{aligned}\int_0^b x^n\mathrm{d}F(x)&=-\int_0^b x^n\mathrm{d}[1-F(x)]\\&=-b^n[1-F(b)]+n\int_0^b x^{n-1}[1-F(x)]\mathrm{d}x.\end{aligned} \tag{72}$$

现在, 在 $\mathbf{E}|\xi|^n<\infty$ 的条件下, 证明,

$$b^n[1-F(b)-F(-b)]\leqslant b^n\mathbf{P}\{|\xi|\geqslant b\}\to 0, b\to\infty. \tag{73}$$

事实上,

$$\mathbf{E}|\xi|^n=\sum_{k=1}^{\infty}\int_{k-1}^{k}|x|^n\mathrm{d}F(x)<\infty,$$

因而

$$\sum_{k\geqslant b+1}\int_{k-1}^{k}|x|^n\mathrm{d}F(x)\to 0, b\to\infty.$$

由于

$$\sum_{k \geqslant b+1} \int_{k-1}^{k} |x|^n \mathrm{d}F(x) \geqslant b^n \mathbf{P}\{|\xi| \geqslant b\},$$

可见 (73) 式成立.

在 (72) 式中令 $b \to \infty$ 求极限, 得欲证明的 (69) 式. 类似可以证明 (70) 式.

由 (69) 和 (70) 式得公式 (71).

13. 有界变差函数 设 $A = A(t), t \geqslant 0$ 是右连续、有左极限且局部有界变差函数 (即在每一个有限区间 $[a, b]$ 上有有界变差). 考虑方程

$$Z_t = 1 + \int_0^t Z_{s-} \mathrm{d}A(s), \tag{74}$$

其微分形式为

$$\mathrm{d}Z_t = Z_{t-} \mathrm{d}A(t), \quad Z_0 = 1. \tag{75}$$

利用上一小节证明的分部积分公式, 可以 (在局部有界变差函数类中) 求出方程 (74) 之解的明显形式.

引进函数 (称做*随机分量*; [87]):

$$\mathscr{E}_t(A) = \mathrm{e}^{A(t)-A(0)} \prod_{0<s\leqslant t} (1 + \Delta A(s)) \mathrm{e}^{-\Delta A(s)}, \tag{76}$$

其中 $\Delta A(s) = A(s) - A(s-), s > 0$, 而 $A(0) = 0$.

函数 $A(s), 0 \leqslant s \leqslant t$ 具有有界变差, 故最多有可数个间断点, 而且级数 $\sum_{0<s\leqslant t} |A(s)|$ 收敛. 由此可见函数

$$\prod_{0<s\leqslant t} (1 + \Delta A(s)) \mathrm{e}^{-\Delta A(s)}, \quad t \geqslant 0,$$

是局部有界变差函数.

如果记

$$A^c(t) = A(t) - \sum_{0<s\leqslant t} A(s)$$

为函数 $A(t)$ 的连续分量, 则 (76) 式可以改写成如下形式:

$$\mathscr{E}_t(A) = \mathrm{e}^{A^c(t)-A^c(0)} \prod_{0<s\leqslant t} (1 + \Delta A(s)). \tag{77}$$

记

$$F(t) = \mathrm{e}^{A^c(t)-A^c(0)}, G(t) = \prod_{0<s\leqslant t} (1 + \Delta A(s)), \quad G(0) = 1,$$

那么, 由 (62) 式, 有

$$\mathscr{E}_t(A) = F(t)G(t) = 1 + \int_0^t F(s)\mathrm{d}G(s) + \int_0^t G(s-)\mathrm{d}F(s)$$
$$= 1 + \sum_{0<s\leqslant t} F(s)G(s-)\Delta A(s) + \int_0^t G(s-)F(s)\mathrm{d}A^c(s)$$
$$= 1 + \int_0^t \mathscr{E}_{s-}(A)\mathrm{d}A(s).$$

于是, $\mathscr{E}_t(A), t \geqslant 0$ 是方程 (74) 的 (局部有界) 解.

现在证明, 此解在局部有界类中唯一.

假设有两个局部有界解: 而 $Y = Y(t), t \geqslant 0$ 是二者之差, 那么

$$Y(t) = \int_0^t Y(s-)\mathrm{d}A(s).$$

设

$$T = \inf\{t \geqslant 0 : Y(t) \neq 0\},$$

如果对于一切 $t \geqslant 0, Y(t) = 0$, 则认为 $T = \infty$.

由于 $A(t), t \geqslant 0$ 是局部有界变差函数, 则存在两个广义分布函数 $A_1(t)$ 和 $A_2(t)$, 使 $A(t) = A_1(t) - A_2(t)$. 若设 $T < \infty$, 则存在有限 $T' > T$, 使

$$[A_1(T') + A_2(T')] - [A_1(T) + A_2(T)] \leqslant \frac{1}{2}.$$

那么, 由方程

$$Y(t) = \int_T^t Y(s-)\mathrm{d}A(s), t \geqslant T,$$

可见

$$\sup_{t\leqslant T'} |Y(t)| \leqslant \frac{1}{2} \sup_{t\leqslant T'} |Y(t)|.$$

因为 $\sup_{t\leqslant T'} |Y(t)| < \infty$, 所以对于 $T < t \leqslant T', Y(t) = 0$. 而这与假设矛盾.

于是, 证明了下面的定理.

定理 12　*在局部有界函数类中, 方程 (74) 有由 (76) 式给出的解, 而且是唯一解.*

14. 练习题

1. 证明 (6) 式.

2. 证明性质 **E** 的如下推广. 设对于随机变量 ξ 和 η 数学期望 $\mathbf{E}\xi$ 和 $\mathbf{E}\eta$ 存在, 且 $\mathbf{E}\xi + \mathbf{E}\eta$ 有意义 (不是 $\infty - \infty$ 或 $-\infty + \infty$), 则

$$\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta.$$

3. 推广性质 **G**: 证明, 如果 $\xi = \eta$ (a.c.), 且 $\mathbf{E}\xi$ 存在, 则 $\mathbf{E}\eta$ 也存在, 并且 $\mathbf{E}\xi = \mathbf{E}\eta$.

4. 假设 ξ 是广义随机变量, μ 是 $\sigma-$ 有限测度, 且 $\int_\Omega |\xi| \mathrm{d}\mu < \infty$. 证明 $|\xi| < \infty$ (a. c.). (与性质 **J** 比较.)

5. 设 μ 是 $\sigma-$ 有限测度, ξ 和 η 是广义随机变量, 且 $\int \xi \mathrm{d}\mu$ 和 $\int \eta \mathrm{d}\mu$ 存在. 证明, 如果对于一切 $A \in \mathscr{F}$, 有

$$\int_A \xi \mathrm{d}\mu \leqslant \int_A \eta \mathrm{d}\mu,$$

则 $\xi \leqslant \eta$ ($\mu-$a.c.). (与性质 **I** 比较).

6. 设 ξ 和 η 是独立非负随机变量, 证明 $\mathbf{E}\xi\eta = \mathbf{E}\xi \times \mathbf{E}\eta$.

7. 利用法图引理, 证明

$$\mathbf{P}(\underline{\lim} A_n) \leqslant \underline{\lim} \mathbf{P}(A_n), \quad \mathbf{P}(\overline{\lim} A_n) \leqslant \overline{\lim} \mathbf{P}(A_n).$$

8. 举例说明, 控制收敛定理的条件 “$|\xi_n| \leqslant \eta, \mathbf{E}\eta < \infty$” 一般不能减弱.

9. 举例说明, 法图引理的条件 “$\xi_n \leqslant \eta, \mathbf{E}\eta > -\infty$” 一般不能去掉.

10. 证明法图引理的如下变形: 若随机变量序列 $\{\xi_n^+\}_{n\geqslant 1}$ 一致可积, 则

$$\overline{\lim} \mathbf{E}\xi_n \leqslant \mathbf{E}\overline{\lim}\xi_n.$$

11. 定义在 $[0,1]$ 上的狄利克雷 (P. G. L. Dirichlet) 函数

$$d(x) = \begin{cases} 1, & \text{若 } x \text{ 是无理数}, \\ 0, & \text{若 } x \text{ 是有理数}, \end{cases}$$

勒贝格可积, 但是黎曼不可积. 为什么?

12. 举例说明, 定义在 $[0,1]$ 上且黎曼可积的函数序列 $\{f_n\}_{n\geqslant 1}, |f_n| \leqslant 1$, 且关于勒贝格测度几乎处处有 $f_n \to f$, 但是 f 黎曼不可积.

13. 设 $\{a_{ij}, i, j \geqslant 1\}$ 是实数序列, 且 $\sum_{i,j} |a_{ij}| < \infty$. 由傅比尼定理导出

$$\sum_{(i,j)} a_{ij} = \sum_i \left(\sum_j a_{ij} \right) = \sum_j \left(\sum_i a_{ij} \right). \tag{78}$$

14. 举一数列 $\{a_{ij}, i, j \geqslant 1\}$ 的例子, 使 $\sum_{i,j} |a_{ij}| = \infty$ 而等式 (78) 不成立.

15. 从简单函数出发, 利用在积分号下取极限的勒贝格定理, 证明, 关于利用换元积分法的如下结果.

设 $h = h(x)$ 是区间 $[a, b]$ 上的非减连续可微函数, 而 $f(x)$ 在 $[h(a), h(b)]$ 上的 (关于勒贝格测度) 可积. 那么, 函数 $f(h(y))h'(y)$ 在区间 $[a, b]$ 上可积, 且

$$\int_{h(a)}^{h(b)} f(x) \mathrm{d}x = \int_a^b f(h(y))h'(y) \mathrm{d}y.$$

16. 证明 (70) 式.

17. 设 $\xi,\xi_1,\xi_2,\cdots$ 是非负可积随机变量, 满足 $\mathbf{E}\xi_n\to\mathbf{E}\xi$, 且对于任意 $\varepsilon>0$, 概率 $\mathbf{P}\{|\xi-\xi_n|>\varepsilon\}\to 0$. 证明 $\mathbf{E}|\xi_n-\xi|\to 0, n\to\infty$.

18. 设 ξ 是可积随机变量 ($\mathbf{E}|\xi|<\infty$). 证明, 对于任意 $\varepsilon>0$, 存在 $\delta>0$, 使对于任何 $A\in\mathscr{F},\mathbf{P}(A)<\delta$, 有 $\mathbf{E}I_A|\xi|<\varepsilon$ ("勒贝格积分的绝对连续性").

19. 设对于随机变量 ξ,η,ζ 和 $\xi_n,\eta_n,\zeta_n,n\geqslant 1$, 有*)

$$\xi_n\xrightarrow{\mathbf{P}}\xi,\eta_n\xrightarrow{\mathbf{P}}\eta,\zeta_n\xrightarrow{\mathbf{P}}\zeta,\eta_n\leqslant\xi_n\leqslant\zeta_n,n\geqslant 1,$$

$$\mathbf{E}\zeta_n\longrightarrow\mathbf{E}\zeta,\mathbf{E}\eta_n\longrightarrow\mathbf{E}\eta,$$

且数学期望 $\mathbf{E}\xi,\mathbf{E}\eta,\mathbf{E}\zeta$ 有限. 证明普拉特 (Pratt) 引理: $\mathbf{E}\xi_n\to\mathbf{E}\xi$, 而如果满足 $\eta_n\leqslant 0\leqslant\zeta_n$, 则 $\mathbf{E}|\xi_n-\xi|\to 0$.

由此证明, 如果 $\xi_n\xrightarrow{\mathbf{P}}\xi,\mathbf{E}|\xi_n|\to\mathbf{E}|\xi|$ 和 $\mathbf{E}|\xi|<\infty$, 则 $\mathbf{E}|\xi_n-\xi|\to 0$.

举例说明, 在普拉特引理的条件下, 一般 $\mathbf{E}|\xi_n-\xi|\not\to 0$.

20. 证明 $L_*f\leqslant L^*f$, 而如果函数 f 有界, 且测度 μ 有限, 则 $L_*f=L^*f$ (见第 11 小节之注 2).

21. 证明, 对于有界函数 f, 数学期望 $\mathbf{E}f=L_*f$ (见第 11 小节之注 2).

22. 证明第 11 小节之注 2 的最后一个命题.

23. 设 $F(x)$ 是随机变量 X 的分布函数. 证明

$$\mathbf{E}X^+<\infty\Leftrightarrow\int_a^\infty\ln\frac{1}{F(x)}\mathrm{d}x<\infty,\ \text{对于某个}\ a.$$

24. 证明, 如果对于 $p>0$,

$$\lim_{x\to\infty}x^p\mathbf{P}\{|\xi|>x\}=0,$$

则对于一切 $r<p,\mathbf{E}|\xi|^r<\infty$. 举例说明, 当 $r=p$ 时有可能 $\mathbf{E}|\xi|^p=\infty$.

25. 举一密度 $f(x)$ 的例子, $f(x)$ 是偶函数, 然而所有非奇数阶矩

$$\int_{-\infty}^\infty x^kf(x)\mathrm{d}x=0,\quad k=1,3,\cdots$$

26. 举一随机变量序列 $\xi_n,n\geqslant 1$ 的例子, 使

$$\mathbf{E}\sum_{n=1}^\infty\xi_n\neq\sum_{n=1}^\infty\mathbf{E}\xi_n.$$

27. 设对于随机变量 X, 当 $\alpha>1$ 时, 有

$$\frac{\mathbf{P}\{|X|>\alpha n\}}{\mathbf{P}\{|X|>n\}}\to 0,n\to\infty.$$

*) 收敛 $\xi_n\xrightarrow{\mathbf{P}}\xi$ 称做依概率收敛, 表示对于任意 $\varepsilon>0$, 概率 $\mathbf{P}\{|\xi_n-\xi|>\varepsilon\}\to 0,n\to\infty$. 详见 §10.

证明 X 一切阶矩存在. **提示**: 利用公式:

$$\mathbf{E}|X|^N = N\int_0^\infty x^{N-1}\mathbf{P}\{|X|>x\}\mathrm{d}x.$$

28. 设对于随机变量 X 以概率 p_k 取 $k=0,1,2,\cdots$ 为值. 函数

$$F(s)=\sum_{k=0}^{\infty}p_k s^k,\quad |s|\leqslant 1,$$

称做随机变量 X 的**母函数**. 证明下列公式:

(i) 若 X 是泊松随机变量, 即 $p_k=\mathrm{e}^{-\lambda}\lambda^k/k!$, 其中 $\lambda>0, k=0,1,2,\cdots$, 则

$$F(s)=\mathrm{e}^{-\lambda(1-s)}, |s|\leqslant 1;$$

(ii) 若随机变量 X 有几何分布, 即 $p_k=pq^k$, 其中 $0<p<1, q=1-p, k=0,1,2,\cdots$, 则

$$F(s)=\frac{p}{1-sq},\quad |s|\leqslant 1.$$

29. 除母函数 $F(s)$ 外尚使用**矩母函数**: $M(s)=\mathbf{E}\mathrm{e}^{sX}$ (假设 s 使 $\mathbf{E}\mathrm{e}^{sX}<\infty$).

(a) 证明, 若矩母函数 $M(s)$ 对于 0 的某一邻域的一切 $s(s\in[-a,a], a>0)$ 有定义, 则当 $s=0$ 时对于所有 $k=1,2,\cdots$, 导数 $M^{(k)}(s)$ 存在, 而且

$$M^{(k)}(0)=\mathbf{E}X^k$$

(这一性质决定了 $M(s)$ 的名称).

(b) 举一随机变量的例子, 对于一切 $s>0$, 其矩母函数 $M(s)=\infty$.

(c) 证明, 对于一切 $s\in\mathbb{R}$, 参数 $\lambda>0$ 的泊松随机变量的矩母函数为 $M(s)=\mathrm{e}^{-\lambda(1-\mathrm{e}^s)}$.

30. 设 $0<r<\infty, X_n\in L^r, X_n\xrightarrow{\mathbf{P}}X$. 证明下列条件等价:

(i) 随机变量族 $\{|X_n|^r, n\geqslant 1\}$ 一致可积;

(ii) 在 L^r 中 $X_n\to X$;

(iii) $\mathbf{E}|X_n|^r\to\mathbf{E}|X|^r<\infty$.

31. **斯皮策** (F. Spitzer) **恒等式**. 设 $X_1, X_2,\cdots$ 是独立同分布随机变量序列, $\mathbf{P}\{X_1\leqslant 1\}=1$; $S_n=X_1+\cdots+X_n$, 则对于 $|u|,|t|<1$, 有

$$\sum_{n=0}^{\infty}t^n\mathbf{E}\mathrm{e}^{uM_n}=\exp\left\{\sum_{n=1}^{\infty}\frac{t^n}{n}\mathbf{E}(uS_n^u)\right\},$$

其中 $M_n=\max\{0,X_1,X_2,\cdots,X_n\}, S_n^u=\max\{u,S_n\}$.

32. 设 $S_0=0, S_n=X_1+X_2+\cdots+X_n, n\geqslant 1$ 是简单对称随机游动, 而 $\tau=\min\{n>0:S_n\geqslant 0\}$. 证明

$$\mathbf{E}\min(\tau,2m)=2\mathbf{E}|S_{2m}|=4m\mathbf{P}\{S_{2m}=0\}, m\geqslant 0.$$

33. 设 ξ 是标准正态随机变量: $\xi \sim N(0,1)$. 利用分部积分法证明 $\mathbf{E}\xi^k = (k-1)\mathbf{E}\xi^{k-2}$, 并由此导出公式:

$$\mathbf{E}\xi^{2k-1} = 0 \text{ 和 } \mathbf{E}\xi^{2k} = 1 \times 3 \times \cdots \times (2k-3)(2k-1) \quad (= (2k-1)!!).$$

34. 证明, 函数 $x^{-1} \sin x, x \in \mathbb{R}$ 按黎曼可积, 但是按 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的勒贝格测度不可积.

35. 证明函数

$$\xi(\omega_1, \omega_2) = \mathrm{e}^{-\omega_1\omega_2} - 2\mathrm{e}^{-2\omega_1\omega_2}, \quad \omega_1 \in \Omega_1 = [1, \infty), \quad \omega_2 \in \Omega_2 = (0, 1],$$

关于勒贝格测度,

(a) 对于每一个 ω_2, 对 $\omega_1 \in \Omega_1$ 可积;

(b) 对于每一个 ω_1, 对 $\omega_2 \in \Omega_2$ 可积, 但是傅比尼定理不成立.

36. 证明莱维 (P. P. Lévy) 定理: 设随机变量 $\xi_1, \xi_2, \cdots$ 可积 (对于一切 $n \geqslant 1, \mathbf{E}|\xi_n| < \infty$), $\sup_n \mathbf{E}\xi_n < \infty$ 且 $\xi_n \uparrow \xi$, 则随机变量 ξ 可积, 且 $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi$ (对照定理 1 之 a)).

37. 证明法图引理的如下变式: 如果 $0 \leqslant \xi_n \to \xi$ (P–a.c.), 且 $\mathbf{E}\xi_n \leqslant A < \infty$, 则 ξ 可积, 且 $\mathbf{E}\xi \leqslant A$.

38. (黎曼积分法与勒贝格积分法的联系.) 设博雷尔函数 $f = f(x)$ 对勒贝格测度可积: $\int_{\mathbb{R}} |f(x)| \mathrm{d}x < \infty$. 证明对于任意 $\varepsilon > 0$, 存在

(a) 具有有界区间 A_i 的阶梯函数

$$f_\varepsilon(x) = \sum_{i=1}^{n} f_i I_{A_i}(x), \quad \text{使} \quad \int_{\mathbb{R}} |f(x) - f_\varepsilon(x)| \mathrm{d}x < \varepsilon;$$

(b) 具有有界承载子的连续可积函数 $g_\varepsilon(x)$, 使

$$\int_{\mathbb{R}} |f(x) - g_\varepsilon(x)| \mathrm{d}x < \varepsilon.$$

39. 证明, 如果 ξ 是可积随机变量, 则

$$\mathbf{E}\xi = \int_0^\infty \mathbf{P}\{\xi > x\} \mathrm{d}x - \int_{-\infty}^0 \mathbf{P}\{\xi < x\} \mathrm{d}x.$$

40. 证明, 如果 ξ 和 η 是可积随机变量, 则

$$\mathbf{E}\xi - \mathbf{E}\eta = \int_{-\infty}^{\infty} [\mathbf{P}\{\eta < x \leqslant \xi\} - \mathbf{P}\{\xi < x \leqslant \eta\}] \mathrm{d}x.$$

41. 设 ξ 是非负随机变量 ($\xi \geqslant 0$), 其拉普拉斯变换为 $\varphi_\xi(\lambda) = \mathbf{E}\mathrm{e}^{-\lambda\xi}, \lambda \geqslant 0$.

(a) 证明, 对于任意 $0 < r < 1$,

$$\mathbf{E}\xi^r = \frac{r}{\Gamma(1-r)} \int_0^\infty \frac{1 - \varphi_\xi(\lambda)}{\lambda^{r+1}} \mathrm{d}\lambda.$$

提示 对于 $s \geqslant 0, 0 < r < 1$,

$$\frac{1}{r}\Gamma(1-r)s^r = \int_0^\infty \frac{1-\mathrm{e}^{-s\lambda}}{\lambda^{r+1}}\mathrm{d}\lambda.$$

(b) 证明, 如果 $\xi > 0$, 则对于任意 $r > 0$,

$$\mathbf{E}\xi^{-r} = \frac{1}{r\Gamma(r)}\int_0^\infty \varphi_\xi(\lambda^{1/r})\mathrm{d}\lambda.$$

提示 对于 $s \geqslant 0, r > 0$,

$$s = \frac{r}{\Gamma(1/r)} = \int_0^\infty \exp\{-(\lambda/s)^r\}\mathrm{d}\lambda.$$

§7. 关于 $\sigma-$ 代数的条件概率和条件数学期望

1. 条件数学期望和条件概率一般定义的必要性 设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, 而事件 $A \in \mathscr{F}$ 的概率 $\mathbf{P}(A) > 0$;$\mathscr{D} = \{D_1, D_2, \cdots\}$ 是空间 Ω 的有限或可数分割, 其中 $\mathbf{P}(D_i) > 0, i \geqslant 1$. 像有限概率空间一样, 称

$$\mathbf{P}(B|A) \equiv \frac{\mathbf{P}(AB)}{\mathbf{P}(A)}$$

为事件 B 关于事件 A 的条件概率. 而事件 B 关于分割 $\mathscr{D}$ 的条件概率, 记作 $\mathbf{P}(B|\mathscr{D})$ 或 $\mathbf{P}(B|\mathscr{D})(\omega)$, 定义为: 对于 $\omega \in D_i, i \geqslant 1$, 等于 $\mathbf{P}(B|D_i)$ 的随机变量

$$\mathbf{P}(B|\mathscr{D})(\omega) = \sum_{i\geqslant 1}\mathbf{P}(B|D_i)I_{D_i}(\omega).$$

类似地, 如果 ξ 是随机变量, 假设其数学期望 $\mathbf{E}\xi$ 存在, 则 ξ 关于事件 $A(\mathbf{P}(A) > 0)$ *的条件数学期望*定义为

$$\mathbf{E}(\xi|A) \equiv \frac{\mathbf{E}(\xi I_A)}{\mathbf{P}(A)}$$

(对照第一章 §8 的 (10) 式).

随机变量 $\mathbf{P}(B|\mathscr{D})$ 显然关于 $\sigma-$ 代数 $\mathscr{G} = \sigma(\mathscr{D})$ 可测, 因此 $\mathbf{P}(B|\mathscr{D})$ 亦记作 $\mathbf{P}(B|\mathscr{G})$(见第一章 §8).

在概率论中往往有必要考虑关于零概率事件的条件概率.

例如, 考虑如下试验. 设 ξ 是在 $[0,1]$ 上均匀分布的随机变量. 如果 $\xi = x$, 则掷一次硬币:正面出现的概率为 x, 反面出现的概率为 $1-x$. 以 ν 表示将这样的硬币掷 n 次, 正面出现的次数. 问 "条件概率 $\mathbf{P}\{\nu = k|\xi = x\}$ 如何?" 尽管直观上清楚 "这一概率似乎应该等于 $\mathrm{C}_n^k x^k(1-x)^{n-k}$", 然而, 由于当 $\mathbf{P}\{\xi = x\} = 0$ 时, 我们所关心的 "条件概率 $\mathbf{P}\{\nu = k|\xi = x\}$" 暂时没有定义.

下面将给出关于 $\sigma-$ 代数 $\mathscr{G}, \mathscr{G} \subseteq \mathscr{F}$, 条件数学期望 (其中包括条件概率) 的一般定义, 并且与第一章 §8 对于有限概率空间的相应定义作比较.

2. 条件数学期望和条件概率的一般定义 设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, $\mathscr{G}$ 是 $\sigma-$ 代数, $\mathscr{G} \subseteq \mathscr{F}$($\mathscr{G}$ 是 $\mathscr{F}$ 的子 $\sigma-$ 代数), 而 $\xi = \xi(\omega)$ 是随机变量. 回忆在 §6 中分两个阶段引进了数学期望 $\mathbf{E}\xi$ 的定义: 首先对非负随机变量 ξ, 然后对一般随机变量 ξ, 并利用等式

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^-,$$

其中 (为避免形如 $\infty - \infty$ 的不定式) 仅假设

$$\min\{\mathbf{E}\xi^+, \mathbf{E}\xi^-\} < \infty.$$

在定义条件数学期望 $\mathbf{E}(\xi|\mathscr{G})$ 时, 也采用类似的两阶段构造过程.

定义 1 1) 非负随机变量 ξ 关于 $\sigma-$ 代数 $\mathscr{G}$ 的条件数学期望, 定义为非负 (广义) 随机变量: $\mathbf{E}(\xi|\mathscr{G})$ 或 $\mathbf{E}(\xi|\mathscr{G})(\omega)$, 如果

a) $\mathbf{E}(\xi|\mathscr{G})$ 为 $\mathscr{G}-$ 可测;

b) 对于任意 $A \in \mathscr{G}$, 有

$$\int_A \xi \mathrm{d}\mathbf{P} = \int_A \mathbf{E}(\xi|\mathscr{G})\mathrm{d}\mathbf{P}. \tag{1}$$

2) 假设任意随机变量 ξ 关于 $\sigma-$ 代数 $\mathscr{G}$ 的条件数学期望 $\mathbf{E}(\xi|\mathscr{G})$或 $\mathbf{E}(\xi|\mathscr{G})(\omega)$ 有定义, 如果 $\mathbf{P}-$a.c.

$$\min\{\mathbf{E}(\xi^+|\mathscr{G}), \quad \mathbf{E}(\xi^-|\mathscr{G})\} < \infty,$$

并且通过如下公式表示

$$\mathbf{E}(\xi|\mathscr{G}) \equiv \mathbf{E}(\xi^+|\mathscr{G}) - \mathbf{E}(\xi^-|\mathscr{G}),$$

且在使 $\mathbf{E}(\xi^+|\mathscr{G}) = \mathbf{E}(\xi^-|\mathscr{G}) = \infty$ 的一切基本事件的 (零概率) 集合上, 任意规定差 $\mathbf{E}(\xi^+|\mathscr{G}) - \mathbf{E}(\xi^-|\mathscr{G})$, 例如令其等于 0.

首先证明, 对于非负随机变量 ξ,$\mathbf{E}(\xi|\mathscr{G})$ 确实存在. 根据 §6 第 8 小节, 集函数

$$\mathbf{Q}(A) = \int_A \xi \mathrm{d}\mathbf{P}, \ A \in \mathscr{G} \tag{2}$$

是在 $(\Omega, \mathscr{G})$ 上关于测度 $\mathbf{P}$ 绝对连续的测度, 其中 $\mathbf{P}$ 是在 $(\Omega, \mathscr{G}), \mathscr{G} \subseteq \mathscr{F}$ 上的测度. 因此 (根据拉东 – 尼科迪姆定理) 存在这样的一个 $\mathscr{G}-$ 可测广义随机变量 $\mathbf{E}(\xi|\mathscr{G})$, 使

$$\mathbf{Q}(A) = \int_A \mathbf{E}(\xi|\mathscr{G})\mathrm{d}\mathbf{P}. \tag{3}$$

由 (2) 和 (3) 式得 (1) 式.

注 1 根据拉东 – 尼科迪姆定理, 条件数学期望 $\mathbf{E}(\xi|\mathscr{G})$ 仅精确到 $\mathbf{P}-$ 零测集唯一. 换句话说, 作为 $\mathbf{E}(\xi|\mathscr{G})(\omega)$ 可以取任何 $\mathscr{G}-$ 可测函数 $f(\omega)$, 其中 $f(\omega)$ 满足

$$\mathbf{Q}(A) = \int_A f(\omega)\mathrm{d}\mathbf{P}, A \in \mathscr{G},$$

称做条件数学期望的变式.

还应指出, 根据对拉东 - 尼科迪姆定理的注释,

$$\mathbf{E}(\xi|\mathscr{G}) \equiv \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}(\omega), \tag{4}$$

即条件数学期望是测度 $\mathbf{Q}$ 对测度 $\mathbf{P}$ 的拉东 - 尼科迪姆导数, 其中测度 $\mathbf{Q}$ 和测度 $\mathbf{P}$ 都是空间 $(\Omega,\mathscr{G})$ 上的测度.

有益地指出, 由 (1) 式可见, 如果对于非负随机变量 $\xi,\mathbf{E}\xi<\infty$, 则 $\mathbf{E}(\xi|\mathscr{G})<\infty$ ($\mathbf{P}$–a.c.). 类似地, 如果 $\xi\leqslant 0,\mathbf{E}\xi>-\infty$, 则 $\mathbf{E}(\xi|\mathscr{G})>-\infty$ ($\mathbf{P}$–a.c.).

注 2 关于 (1) 需要指出, 一般不能设 $\mathbf{E}(\xi|\mathscr{G})=\xi$, 因为随机变量 ξ 未必 $\mathscr{G}-$ 可测.

注 3 假设对于随机变量 $\xi,\mathbf{E}\xi$ 存在, 则 $\mathbf{E}(\xi|\mathscr{G})$ 也许可以定义为使 (1) 式成立的 $\mathscr{G}-$ 可测函数. 通常正是这样做的. 我们引进的定义 $\mathbf{E}(\xi|\mathscr{G})\equiv\mathbf{E}(\xi^+|\mathscr{G})-\mathbf{E}(\xi^-|\mathscr{G})$ 具有如下优点: 在 $\mathscr{G}$ 是平凡 $\sigma-$ 代数的情形下, 即在 $\mathscr{G}=(\varnothing,\Omega)$ 的情形下, $\mathbf{E}(\xi|\mathscr{G})=\mathbf{E}\xi$, 这时并不假定 $\mathbf{E}\xi$ 存在. (例如, 对于随机变量 ξ, 假如 $\mathbf{E}\xi^+=\infty,\mathbf{E}\xi^-=\infty,\mathscr{G}=\mathscr{F}$, 则 $\mathbf{E}\xi$ 没有定义, 但是在按定义 1, $\mathbf{E}(\xi|\mathscr{G})$ 存在, 并且 $\xi=\xi^+-\xi^-$.)

注 4 假设条件数学期望 $\mathbf{E}(\xi|\mathscr{G})$ 有定义. 随机变量

$$\mathbf{D}(\xi|\mathscr{G})=\mathbf{E}\{[\xi-\mathbf{E}(\xi|\mathscr{G})]^2|\mathscr{G}\}$$

称做随机变量 ξ *关于 $\sigma-$ 代数 $\mathscr{G}$ 的条件方差*. (与第一章 §8 练习题 2 中 $\mathbf{D}(\xi|\mathscr{D})$ 关于 $\mathscr{D}$ 的定义, 以及 §8 中方差的定义比较.)

定义 2 设 $B\in\mathscr{F}$. 条件数学期望 $\mathbf{E}(I_B|\mathscr{G})$ 记作 $\mathbf{P}(B|\mathscr{G})$ 或 $\mathbf{P}(B|\mathscr{G})(\omega)$, 并且称做*事件 B 关于 $\sigma-$ 代数 $\mathscr{G}(\mathscr{G}\subseteq\mathscr{F})$ 的条件概率*.

由定义 1 和 2, 对于每一个固定的 $B\in\mathscr{F}$, 条件概率 $\mathbf{P}(B|\mathscr{G})$ 是一随机变量, 满足:

a) $\mathbf{P}(B|\mathscr{G})$ 为 $\mathscr{G}-$ 可测;

b) 对于任意 $A\in\mathscr{G}$,

$$\mathbf{P}(A\cap B)=\int_A\mathbf{P}(B|\mathscr{G})\mathrm{d}\mathbf{P}. \tag{5}$$

定义 3 设 ξ 是随机变量, $\mathscr{G}_\eta$ 是某随机元素 η 诱导的 $\sigma-$ 代数. 那么, 如果 $\mathbf{E}(\xi|\mathscr{G}_\eta)$ 有定义, 则记作 $\mathbf{E}(\xi|\eta)$ 或 $\mathbf{E}(\xi|\eta)(\omega)$, 并称做 *$\xi$ 关于 η 的条件数学期望*.

条件概率 $\mathbf{P}(B|\mathscr{G}_\eta)$, 记作 $\mathbf{P}(B|\eta)$ 或 $\mathbf{P}(B|\eta)(\omega)$, 称做*事件 B 关于 η 的条件概率*.

3. 关于分割和关于 $\sigma-$ 代数的条件数学期望的关系 现在证明, 这里给出的 $\mathbf{E}(\xi|\mathscr{G})$ 的定义与第一章 §8 中条件数学期望的定义一致.

设 $\mathscr{D}=\{D_1,D_2,\cdots\}$ 是以 $D_i(\sum_i D_i=\Omega)$ 为*原子*的有限或可数分割, 其中 $\mathbf{P}(D_i)>0,i\geqslant 1$.

定理 1 如果 $\mathscr{G}=\sigma(\mathscr{D})$, 而 ξ 是随机变量, 且其数学期望 $\mathbf{E}\xi$ 有定义, 则

$$\mathbf{E}(\xi|\mathscr{G})=\mathbf{E}(\xi|D_i),\quad (\text{在 } D_i \text{ 上 } \mathbf{P}-\text{a.c.}) \tag{6}$$

或

$$\mathbf{E}(\xi|\mathscr{G})=\frac{\mathbf{E}(\xi I_{D_i})}{\mathbf{P}(D_i)},\quad (\text{在 } D_i \text{ 上 } \mathbf{P}-\text{a.c.}).$$

(其中 "$\xi=\eta$ (在 A 上 $\mathbf{P}$–a.c.)" 或 "$\xi=\eta$ (A;$\mathbf{P}$–a.c.)" 表示 $\mathbf{P}(A\cap\{\xi\neq\eta\})=0$.)

证明 根据 §4 引理 3, 在 D_i 上 $\mathbf{E}(\xi|\mathscr{G})=K_i$, 其中 K_i 是常数. 由

$$\int_{D_i}\xi\mathrm{d}\mathbf{P}=\int_{D_i}\mathbf{E}(\xi|\mathscr{G})\mathrm{d}\mathbf{P}=K_i\mathbf{P}(D_i),$$

可见

$$K_i=\frac{1}{\mathbf{P}(D_i)}\int_{D_i}\xi\mathrm{d}\mathbf{P}=\frac{\mathbf{E}(\xi I_{D_i})}{\mathbf{P}(D_i)}=\mathbf{E}(\xi|D_i).$$ □

于是, 第一章引进的、关于分割 $\mathscr{D}=\{D_1,D_2,\cdots\}$ 的条件数学期望 $\mathbf{E}(\xi|\mathscr{G})$ 的概念, 是关于 $\sigma-$ 代数 $\mathscr{G}=\sigma(\mathscr{D})$ 的条件数学期望概念的特殊情形.

4. 条件数学期望的性质 假设下面所考虑的随机变量 ξ,η 的条件数学期望都存在, 而 $\sigma-$ 代数 $\mathscr{G}\subseteq\mathscr{F}$.

A*. 若 C 是常数, 而 $\xi=C$ (a.c.), 则 $\mathbf{E}(\xi|\mathscr{G})=C$ (a.c.).

B*. 若 $\xi\leqslant\eta$ (a.c.), 则 $\mathbf{E}(\xi|\mathscr{G})\leqslant\mathbf{E}(\eta|\mathscr{G})$ (a.c.).

C*. $|\mathbf{E}(\xi|\mathscr{G})|\leqslant\mathbf{E}(|\xi||\mathscr{G})$ (a. c.).

D*. 若 a,b 为常数, 且 $a\mathbf{E}\xi+b\mathbf{E}\eta$ 存在, 则

$$\mathbf{E}(a\xi+b\eta|\mathscr{G})=a\mathbf{E}(\xi|\mathscr{G})+b\mathbf{E}(\eta|\mathscr{G})\ \text{(a.c.)}.$$

E*. 若 $\mathscr{F}_*=\{\varnothing,\Omega\}$ 是平凡 $\sigma-$ 代数, 则

$$\mathbf{E}(\xi|\mathscr{F}_*)=\mathbf{E}\xi\ \text{(a.c.)}.$$

F*. $\mathbf{E}(\xi|\mathscr{F})=\xi$ (a.c.).

G*. $\mathbf{E}[\mathbf{E}(\xi|\mathscr{G})]=\mathbf{E}\xi$.

H*. 若 $\mathscr{G}_1\subseteq\mathscr{G}_2$, 则有 (第一) "望远性":

$$\mathbf{E}[\mathbf{E}(\xi|\mathscr{G}_2)|\mathscr{G}_1]=\mathbf{E}(\xi|\mathscr{G}_1)\ \text{(a.c.)}.$$

I*. 若 $\mathscr{G}_1\supseteq\mathscr{G}_2$, 则有 (第二) "望远性":

$$\mathbf{E}[\mathbf{E}(\xi|\mathscr{G}_2)|\mathscr{G}_1]=\mathbf{E}(\xi|\mathscr{G}_2)\ \text{(a.c.)}.$$

J*. 若随机变量 ξ 的数学期望 $\mathbf{E}\xi$ 存在, 且与 $\sigma-$ 代数 $\mathscr{G}$ 无关 (即不依赖于 $I_B, B\in\mathscr{G}$), 则

$$\mathbf{E}(\xi|\mathscr{G})=\mathbf{E}\xi\ \text{(a.c.)}.$$

K*. 若 η 是 $\mathscr{G}-$ 可测随机变量, $\mathbf{E}|\xi|<\infty, \mathbf{E}|\xi\eta|<\infty$, 则

$$\mathbf{E}(\xi\eta|\mathscr{G})=\eta\mathbf{E}(\xi|\mathscr{G}) \text{ (a.c.)}.$$

数学期望性质的证明

a*. 等于常数的函数关于 $\mathscr{G}$ 可测, 因此只需验证等式①;

$$\int_A \xi \mathrm{d}\mathbf{P}=\int_A C \mathrm{d}\mathbf{P}, A \in \mathscr{G}.$$

由于假设 $\xi=C$ (a.c.), 则由 §6 第 3 小节的性质 **G**, 知上面的等式显然.

b*. 若 $\xi \leqslant \eta$ (a.c.), 则由 §6 第 3 小节的性质 **B**, 有

$$\int_A \xi \mathrm{d}\mathbf{P} \leqslant \int_A \eta \mathrm{d}\mathbf{P}, A \in \mathscr{G},$$

从而, 有

$$\int_A \mathbf{E}(\xi|\mathscr{G}) \mathrm{d}\mathbf{P} \leqslant \int_A \mathbf{E}(\eta|\mathscr{G}) \mathrm{d}\mathbf{P}, A \in \mathscr{G}.$$

于是, 由 §6 第 3 小节的性质 **I**, 得所要求的不等式.

c*. 若考虑到明显的不等式 $-|\xi| \leqslant \xi \leqslant|\xi|$, 则由 **B*** 可得所要求的不等式.

d*. 若集合 $A \in \mathscr{G}$, 则由 §6 第 14 小节的练习题 2, 有

$$\begin{aligned}\int_A(a\xi+b\eta) \mathrm{d}\mathbf{P} &=\int_A a\xi \mathrm{d}\mathbf{P}+\int_A b\eta \mathrm{d}\mathbf{P}=\int_A a\mathbf{E}(\xi|\mathscr{G}) \mathrm{d}\mathbf{P}+\int_A b\mathbf{E}(\eta|\mathscr{G}) \mathrm{d}\mathbf{P} \\ &=\int_A[a\mathbf{E}(\xi|\mathscr{G})+b\mathbf{E}(\eta|\mathscr{G})] \mathrm{d}\mathbf{P}.\end{aligned}$$

于是性质 **D*** 得证.

e*. 该性质可以由下列性质得出: 第一, $\mathbf{E}\xi$ 是 $\mathscr{F}_*-$ 可测函数; 第二, 如果 $A=\Omega$ 或 $A=\varnothing$, 则显然

$$\int_A \xi \mathrm{d}\mathbf{P}=\int_A \mathbf{E}\xi \mathrm{d}\mathbf{P}.$$

f*. 由于 ξ 为 $\mathscr{F}-$ 可测, 且

$$\int_A \xi \mathrm{d}\mathbf{P}=\int_A \mathbf{E}(\xi|\mathscr{F}) \mathrm{d}\mathbf{P}, A \in \mathscr{F},$$

则 $\mathbf{E}(\xi|\mathscr{F})=\xi$ (a.c.).

g*. 如果设 $\mathscr{G}_1=(\varnothing, \Omega), \mathscr{G}_2=\mathscr{G}$, 则性质 **G*** 可以由性质 **E*** 和 **H*** 得出.

h*. 设 $A \in \mathscr{G}_1$, 则

$$\int_A \mathbf{E}(\xi|\mathscr{G}_1) \mathrm{d}\mathbf{P}=\int_A \xi \mathrm{d}\mathbf{P}.$$

①小写题号 a* ~ k* 下相应为命题 A* ~ K* 的证明. —— 译者

因为 $\mathscr{G}_1 \subseteq \mathscr{G}_2$, 则 $A \in \mathscr{G}_2$, 因此

$$\int_A \mathbf{E}[\mathbf{E}(\xi|\mathscr{G}_2)|\mathscr{G}_1]\mathrm{d}\mathbf{P} = \int_A \mathbf{E}(\xi|\mathscr{G}_2)\mathrm{d}\mathbf{P} = \int_A \xi \mathrm{d}\mathbf{P}.$$

从而, 对于 $A \in \mathscr{G}_1$, 有

$$\int_A \mathbf{E}(\xi|\mathscr{G}_1)\mathrm{d}\mathbf{P} = \int_A \mathbf{E}[\mathbf{E}(\xi|\mathscr{G}_2)|\mathscr{G}_1]\mathrm{d}\mathbf{P}.$$

仿照 §6 第 3 小节的性质 **I** (亦见 §6 的练习题 5), 可得

$$\mathbf{E}(\xi|\mathscr{G}_1) = \mathbf{E}[\mathbf{E}(\xi|\mathscr{G}_2)|\mathscr{G}_1] \quad \text{(a.c.)}.$$

i*. 设 $A \in \mathscr{G}_1$, 则根据 $\mathbf{E}[\mathbf{E}(\xi|\mathscr{G}_2)|\mathscr{G}_1]$ 的定义

$$\int_A \mathbf{E}[\mathbf{E}(\xi|\mathscr{G}_2)|\mathscr{G}_1]\mathrm{d}\mathbf{P} = \int_A \mathbf{E}(\xi|\mathscr{G}_2)\mathrm{d}\mathbf{P}.$$

因为函数 $\mathbf{E}(\xi|\mathscr{G}_2)$ 为 $\mathscr{G}_2-$ 可测, 并且由于 $\mathscr{G}_2 \subseteq \mathscr{G}_1$, 则 $\mathbf{E}(\xi|\mathscr{G}_2)$ 也 $\mathscr{G}_1-$ 可测. 可见 $\mathbf{E}(\xi|\mathscr{G}_2)$ 是条件数学期望 $\mathbf{E}[\mathbf{E}(\xi|\mathscr{G}_2)|\mathscr{G}_1]$ 的变式之一. 于是, 性质 **I*** 得证.

j*. 由于 $\mathbf{E}\xi$ 为 $\mathscr{G}-$ 可测函数, 故只需验证对于任意 $B \in \mathscr{G}$, 有

$$\int_B \xi \mathrm{d}\mathbf{P} = \int_B \mathbf{E}\xi \mathrm{d}\mathbf{P},$$

即 $\mathbf{E}(\xi I_B) = \mathbf{E}\xi \times \mathbf{E}I_B$. 如果 $\mathbf{E}|\xi| < \infty$, 则这由 §6 的定理 6 立刻得到. 对于一般情形, 不是用 §6 的定理 6, 而是利用 §6 的练习题 6 证明上面的积分等式.

k*. 为证明性质 **K***, 需要利用下面定理 2 的命题 a), 因此下面再证明 (见 232 页).

定理 2 (数学期望符号下的收敛性) *设 $\{\xi_n\}_{n\geqslant 1}$ 是 (广义) 随机变量序列.*

a) 若 $|\xi_n| \leqslant \eta, \mathbf{E}\eta < \infty$ 且 $\xi_n \to \xi$ (a.c.), 则

$$\mathbf{E}(\xi_n|\mathscr{G}) \to \mathbf{E}(\xi|\mathscr{G}) \text{ (a.c.)},$$

且

$$\mathbf{E}(|\xi_n - \xi||\mathscr{G}) \to 0 \text{ (a.c.)}.$$

b) 若 $\xi_n \geqslant \eta, \mathbf{E}\eta > -\infty$ 且 $\xi_n \uparrow \xi$ (a.c.), 则

$$\mathbf{E}(\xi_n|\mathscr{G}) \uparrow \mathbf{E}(\xi|\mathscr{G}) \text{ (a.c.)}.$$

c) 若 $\xi_n \leqslant \eta, \mathbf{E}\eta < \infty$ 且 $\xi_n \downarrow \xi$ (a.c.), 则

$$\mathbf{E}(\xi_n|\mathscr{G}) \downarrow \mathbf{E}(\xi|\mathscr{G}) \text{ (a.c.)}.$$

d) 若 $\xi_n \geqslant \eta, \mathbf{E}\eta > -\infty$, 则

$$\mathbf{E}(\underline{\lim}\xi_n|\mathscr{G}) \leqslant \underline{\lim}\mathbf{E}(\xi_n|\mathscr{G}) \text{ (a.c.)}.$$

e) 若 $\xi_n \leqslant \eta, \mathbf{E}\eta < \infty$, 则

$$\overline{\lim}\mathbf{E}(\xi_n|\mathscr{G}) \leqslant \mathbf{E}(\underline{\lim}\xi_n|\mathscr{G}) \text{ (a.c.)}.$$

f) 若 $\xi_n \geqslant 0$, 则

$$\mathbf{E}\left(\sum \xi_n|\mathscr{G}\right) = \sum \mathbf{E}(\xi_n|\mathscr{G}) \text{ (a.c.)}.$$

证明 a) 设

$$\zeta_n = \sup_{m\geqslant n} |\xi_m - \xi_n|.$$

由于 $\xi_n \to \xi$ (a.c.), 可见 $\zeta_n \downarrow 0$ (a.c.). 数学期望 $\mathbf{E}\xi_n$ 和 $\mathbf{E}\xi$ 有限, 由于性质 $\mathbf{D}^*$ 和 $\mathbf{C}^*$ (a.c.),

$$|\mathbf{E}(\xi_n|\mathscr{G}) - \mathbf{E}(\xi|\mathscr{G})| = |\mathbf{E}(\xi_n - \xi|\mathscr{G})| \leqslant \mathbf{E}(|\xi_n - \xi||\mathscr{G}) \leqslant \mathbf{E}(\zeta_n|\mathscr{G}).$$

由于 $\mathbf{E}(\zeta_{n+1}|\mathscr{G}) \leqslant \mathbf{E}(\zeta_n|\mathscr{G})$ (a.c.), 则存在极限 $h = \lim_n \mathbf{E}(\zeta_n|\mathscr{G})$ (a.c.). 那么,

$$0 \leqslant \int_\Omega h\mathrm{d}\mathbf{P} \leqslant \int_\Omega \mathbf{E}(\zeta_n|\mathscr{G})\mathrm{d}\mathbf{P} = \int_\Omega \zeta_n\mathrm{d}\mathbf{P} \to 0, \quad n \to \infty,$$

其中由于 $0 \leqslant \zeta_n \leqslant 2\eta, \mathbf{E}\eta < \infty$, 可见最后的结果可以由控制收敛定理得出. 从而,

$$\int_\Omega h\mathrm{d}\mathbf{P} = 0,$$

且由 §6 第 3 小节的性质 $\mathbf{H}$, 知 $h = 0$ (a.c.).

b) 设 $\eta \equiv 0$. 由于 $\mathbf{E}(\xi_n|\mathscr{G}) \leqslant \mathbf{E}(\xi_{n+1}|\mathscr{G})$ (a.c.), 则存在极限 $\zeta(\omega) = \lim_n \mathbf{E}(\xi_n|\mathscr{G})$ (a.c.). 那么, 由等式

$$\int_A \xi_n\mathrm{d}\mathbf{P} = \int_A \mathbf{E}(\xi_n|\mathscr{G})\mathrm{d}\mathbf{P}, \quad A \in \mathscr{G},$$

和控制收敛定理, 有

$$\int_A \xi\mathrm{d}\mathbf{P} = \int_A \mathbf{E}(\xi|\mathscr{G})\mathrm{d}\mathbf{P} = \int_A \zeta\mathrm{d}\mathbf{P}, \quad A \in \mathscr{G}.$$

从而, 根据与 §6 第 3 小节的性质 $\mathbf{I}$ 类似的性质和 §6 的练习题 5, 可见 $\xi = \zeta$ (a.c.).

为证明一般情形, 注意到, $0 \leqslant \xi_n^+ \uparrow \xi^+$, 且根据已经证明的

$$\mathbf{E}(\xi_n^+|\mathscr{G}) \uparrow \mathbf{E}(\xi^+|\mathscr{G}) \text{ (a.c.)}. \tag{7}$$

由于 $0 \leqslant \xi_n^- \leqslant \xi^-, \mathbf{E}\xi^- < \infty$, 故由 a) 可见

$$\mathbf{E}(\xi_n^-|\mathscr{G}) \to \mathbf{E}(\xi^-|\mathscr{G}),$$

于是, 由此连同 (7) 式, 命题 b) 得证.

由 b) 得命题 c).

d) 设 $\zeta_n = \inf_{m\geqslant n}\xi_m$, 则 $\zeta_n \uparrow \zeta$, 其中 $\zeta = \underline{\lim}\xi_m$. 根据 b), $\mathbf{E}(\zeta_n|\mathscr{G}) \uparrow \mathbf{E}(\zeta|\mathscr{G})$ (a.c.). 因此

$$\mathbf{E}(\underline{\lim}\xi_n|\mathscr{G}) = \mathbf{E}(\zeta|\mathscr{G}) = \lim_n \mathbf{E}(\zeta_n|\mathscr{G}) = \underline{\lim}_n \mathbf{E}(\zeta_n|\mathscr{G}) \leqslant \underline{\lim}\mathbf{E}(\xi_n|\mathscr{G}) \quad \text{(a.c.)}.$$

由 d) 得命题 e).

f) 如果 $\xi_n \geqslant 0$, 则由命题 $\mathbf{D}^*$, 可见

$$\mathbf{E}\left(\sum_{k=1}^n \xi_k \middle| \mathscr{G}\right) = \sum_{k=1}^n \mathbf{E}(\xi_k|\mathscr{G}) \quad \text{(a.c.)},$$

由此连同 b) 即可证明所要求的结果. □

现在证明命题 $\mathbf{K}^*$. 设 $\eta = I_B, B \in \mathscr{G}$. 那么, 对于任意 $A \in \mathscr{G}$, 有

$$\int_A \xi\eta \mathrm{d}\mathbf{P} = \int_{A\cap B} \xi \mathrm{d}\mathbf{P} = \int_{A\cap B} \mathbf{E}(\xi|\mathscr{G})\mathrm{d}\mathbf{P} = \int_A I_B \mathbf{E}(\xi|\mathscr{G})\mathrm{d}\mathbf{P} = \int_A \eta\mathbf{E}(\xi|\mathscr{G})\mathrm{d}\mathbf{P}.$$

由于勒贝格积分的可加性, 等式

$$\int_A \xi\eta \mathrm{d}\mathbf{P} = \int_A \eta\mathbf{E}(\xi|\mathscr{G})\mathrm{d}\mathbf{P}, A \in \mathscr{G}, \tag{8}$$

对于简单随机变量

$$\eta = \sum_{k=1}^n y_k I_{B_k}, B_k \in \mathscr{G},$$

仍然成立. 所以由 §6 第 3 小节的性质 $\mathbf{I}$, 对于简单随机变量, 有

$$\mathbf{E}(\xi\eta|\mathscr{G}) \to \eta\mathbf{E}(\xi|\mathscr{G}) \quad \text{(a.c.)}. \tag{9}$$

现在假设 η 是任意 $\mathscr{G}-$ 可测随机变量, 而 $\{\eta_n\}_{n\geqslant 1}$ 是 $\mathscr{G}-$ 可测简单随机变量序列, 且 $|\eta_n| \leqslant |\eta|, \eta_n \to \eta$. 那么, 由 (9) 式, 可见

$$\mathbf{E}(\xi\eta_n|\mathscr{G}) = \eta_n\mathbf{E}(\xi|\mathscr{G}) \quad \text{(a.c.)}.$$

显然 $|\xi\eta_n| \leqslant |\xi\eta|$, 其中 $|\xi\eta| < \infty$. 因此, 根据定理 2 的命题 a),

$$\mathbf{E}(\xi\eta_n|\mathscr{G}) \to \mathbf{E}(\xi\eta|\mathscr{G}) \quad \text{(a.c.)}.$$

其次, 因为 $\mathbf{E}|\xi| < \infty$, 则 $\mathbf{E}(\xi|\mathscr{G})$ 有限 (a.c.) (见性质 $\mathbf{C}^*$,§6 第 3 小节的性质 $\mathbf{J}$). 于是 $\eta_n\mathbf{E}(\xi|\mathscr{G}) \to \eta\mathbf{E}(\xi|\mathscr{G})$ (a.c.). (关于 $\mathbf{E}(\xi|G)$ 几乎必然有限的假设是重要的, 因为在 §4 第 4 小节曾约定: $0 \times \infty = 0$, 但是如果 $\eta_n = 1/n, \eta \equiv 0$ 就会出现 $(1/n) \times \infty$ 不趋向 $0 \times \infty = 0$ 的情形).

注 如果仅满足条件: η 为 $\mathscr{G}-$可测, 且 $\mathbf{E}(\xi|\mathscr{G})$有定义, 则命题 $\mathbf{K}^*$ 仍然成立.

5. 条件数学期望 $\mathbf{E}(\xi|\mathscr{G}_\eta)$ 的结构 现在比较详细地讨论条件数学期望 $\mathbf{E}(\xi|\mathscr{G}_\eta)$ 的结构, 其中上面曾约定 $\mathbf{E}(\xi|\mathscr{G}_\eta)$ 也通过 $\mathbf{E}(\xi|\eta)$ 表示.

由于 $\mathbf{E}(\xi|\eta)$ 是 $\mathscr{G}-$ 可测函数, 根据 §4 的定理 3(比较确切地说, 该定理对于广义随机变量明显的变式), 存在定义在 $\overline{\mathbb{R}}$ 上取值于 $\overline{\mathbb{R}}$ 的博雷尔函数 $m=m(x)$, 使对于一切 $\omega\in\Omega$, 有

$$m(\eta(\omega))=\mathbf{E}(\xi|\eta)(\omega). \tag{10}$$

我们用 $\mathbf{E}(\xi|\eta=y)$ 表示函数 $m(y)$, 并称做 *ξ 关于事件 $\{\eta=y\}$ 的条件数学期望*, 或 *ξ 在 $\eta=y$ 条件下的条件数学期望*.

根据定义

$$\int_A \xi \mathrm{d}\mathbf{P}=\int_A \mathbf{E}(\xi|\eta)\mathrm{d}\mathbf{P}=\int_A m(\eta)\mathrm{d}\mathbf{P}, A\in\mathscr{G}_\eta. \tag{11}$$

因此, 根据 §6 (关于勒贝格积分变量替换) 的定理 7, 有

$$\int_{\{\omega:\eta\in B\}} m(\eta)\mathrm{d}\mathbf{P}=\int_B m(y)P_\eta(\mathrm{d}y), B\in\mathscr{B}(\overline{\mathbb{R}}). \tag{12}$$

其中 P_η 是 η 的概率分布. 从而 $m=m(y)$ 是博雷尔函数, 且对于一切 $B\in\mathscr{B}(\mathbb{R})$, 有

$$\int_{\{\omega:\eta\in B\}} \xi\mathrm{d}\mathbf{P}=\int_B m(y)P_\eta(\mathrm{d}y). \tag{13}$$

这说明可以经另一途径定义条件数学期望 $\mathbf{E}(\xi|\eta=y)$.

定义 4 设 ξ 和 η 是随机变量 (也可能是广义的), 且存在 $\mathbf{E}\xi$. 我们称任何满足

$$\int_{\{\omega:\eta\in B\}} \xi\mathrm{d}\mathbf{P}=\int_B m(y)P_\eta(\mathrm{d}y), B\in\mathscr{B}(\overline{\mathbb{R}}), \tag{14}$$

的 $\mathscr{B}(\overline{\mathbb{R}})-$ 可测函数 $m=m(x)$, 为*随机变量 ξ 在 $\eta=y$ 条件下的条件数学期望*.

如果注意到集函数

$$\mathbf{Q}(B)=\int_{\{\omega:\eta\in B\}}\mathrm{d}\mathbf{P}$$

是带符号的测度, 并且关于测度 P_η 绝对连续, 则仍然由拉东 – 尼科迪姆定理可见函数 $m=m(x)$ 存在.

现在假设 $m(y)$ 是按定义 4 的条件数学期望. 那么, 仍利用关于勒贝格积分变量替换的定理, 可得

$$\int_{\{\omega:\eta\in B\}} \xi\mathrm{d}\mathbf{P}=\int_B m(y)P_\eta(\mathrm{d}y)=\int_{\{\omega:\eta\in B\}} m(\eta)\mathrm{d}\mathbf{P}, B\in\mathscr{B}(\overline{\mathbb{R}}).$$

函数 $m(\eta)$ 为 $\mathscr{G}_\eta-$ 可测, 且集合 $\{\omega:\eta\in B\}, B\in\mathscr{B}(\overline{\mathbb{R}})$ 包含 $\mathscr{G}_\eta$ 中所有集合.

由此可见, $m(\eta)$ 是数学期望 $\mathbf{E}(\xi|\eta)$. 于是, 由 $\mathbf{E}(\xi|\eta=y)$ 可以得到 $\mathbf{E}(\xi|\eta)$; 相反, 由 $\mathbf{E}(\xi|\eta)$ 可以得到 $\mathbf{E}(\xi|\eta=y)$.

直观上, 条件数学期望 $m(y) = \mathbf{E}(\xi|\eta = y)$ 是比 $\mathbf{E}(\xi|\eta)$ 更简单明了的对象. 不过, 条件数学期望 $\mathbf{E}(\xi|\eta)$ 作为 $\mathscr{G}_\eta-$ 可测随机变量更便于使用.

应该指出, 前面引进的性质 $\mathbf{A}^* \sim \mathbf{K}^*$, 以及定理 2 论断, 容易移植到条件数学期望 $\mathbf{E}(\xi|\eta = y)$(为此, 只需将 "几乎必然" 换成 "$P_\eta-$ 几乎必然"). 例如, 性质 $\mathbf{K}^*$ 可以转述为: 如果, $\mathbf{E}|\xi| < \infty, \mathbf{E}|\xi f(\eta)| < \infty$, 其中 $f = f(y)$ 是 $\mathscr{B}(\overline{\mathbb{R}})-$ 可测函数, 则

$$\mathbf{E}(\xi f(\eta)|\eta = y) = f(y)\mathbf{E}(\xi|\eta = y)(P_\eta - \text{a.c.}). \tag{15}$$

其次 (对照性质 $\mathbf{J}^*$), 设 ξ 和 η 独立, 则

$$\mathbf{E}(\xi|\eta = y) = \mathbf{E}\xi \quad (P_\eta - \text{a.c.}).$$

还要指出, 如果 $B \in \mathscr{B}(\mathbb{R}^2)$ 且 ξ 和 η 独立, 则

$$\mathbf{E}[I_B(\xi, \eta)|\eta = y] = \mathbf{E}I_B(\xi, y) \quad (P_\eta - \text{a.c.}). \tag{16}$$

如果 $\varphi = \varphi(x, y)$ 是 $\mathscr{B}(\mathbb{R}^2)-$ 可测函数, 使 $\mathbf{E}|\varphi(\xi, \eta)| < \infty$, 则

$$\mathbf{E}[\varphi(\xi, \eta)|\eta = y] = \mathbf{E}\varphi(\xi, \eta) \quad (P_\eta - \text{a.c.}).$$

为证明 (16) 式, 我们指出以下事实. 假如 $B = B_1 \times B_2$. 则为证明 (16) 式, 只需验证:

$$\int_{\{\omega:\eta \in A\}} I_{B_1 \times B_2}(\xi, \eta)\mathbf{P}(\mathrm{d}\omega) = \int_{\{y \in A\}} \mathbf{E}I_{B_1 \times B_2}(\xi, y)P_\eta(\mathrm{d}y).$$

等式左侧是 $\mathbf{P}\{\xi \in B_1, \eta \in A \cap B_2\}$, 右侧是 $\mathbf{P}\{\xi \in B_1\}\mathbf{P}\{\eta \in A \cap B_2\}$, 而由于 ξ 和 η 独立, 可见左、右两侧概率相等. 一般情形的证明, 要运用 §2 中关于单调类的定理 1 (对照傅比尼定理证明中的相应方法).

定义 5　*在 $\eta = y$ 的条件下, $\mathbf{E}(I_A|\eta = y)$ 称做事件 $A \in \mathscr{F}$ 的条件概率, 记作 $\mathbf{P}(A|\eta = y)$.*

显然, $\mathbf{P}(A|\eta = y)$ 也可以定义为满足

$$\mathbf{P}(A \cap \{\eta \in B\}) = \int_B \mathbf{P}(A|\eta = y)P_\eta(\mathrm{d}y), \quad B \in \mathscr{B}(\overline{\mathbb{R}}), \tag{17}$$

的可测函数.

6. 计算条件概率和条件数学期望的例

例 1　设 η 是离散型随机变量:

$$\mathbf{P}\{\eta = y_k\} > 0, \quad \sum_{k=1}^{\infty}\mathbf{P}\{\eta = y_k\} = 1.$$

那么

$$\mathbf{P}(A|\eta = y_k) = \frac{\mathbf{P}(A \cap \{\eta = y_k\})}{\mathbf{P}\{\eta = y_k\}}, \quad k \geqslant 1.$$

对于 $y \notin \{y_1, y_2, \cdots\}$, 条件概率 $\mathbf{P}(A|\eta = y)$ 可以任意地定义, 例如, 令其等于 0.

假设随机变量 ξ 的数学期望 $\mathbf{E}\xi$ 存在, 则

$$\mathbf{E}(\xi|\eta = y_k) = \frac{1}{\mathbf{P}\{\eta = y_k\}} \int_{\{\omega:\eta=y_k\}} \xi \mathrm{d}\mathbf{P}.$$

对于 $y \notin \{y_1, y_2, \cdots\}$, 条件数学期望 $\mathbf{E}(\xi|\eta = y)$ 可以任意地定义, 例如, 令其等于 0.

例 2 设 (ξ, η) 是随机向量, 其分布具有密度 $f_{\xi,\eta}(x, y)$:

$$\mathbf{P}\{(\xi, \eta) \in B\} = \int_B f_{\xi,\eta}(x, y)\mathrm{d}x\mathrm{d}y, \quad B \in \mathscr{B}(\mathbb{R}^2).$$

设 $f_\xi(x)$ 和 $f_\eta(y)$ 是随机变量 ξ 和 η 的概率分布密度 (见 §6 (46),(55),(56)). 记

$$f_{\xi|\eta}(x|y) = \frac{f_{\xi,\eta}(x, y)}{f_\eta(y)}, \tag{18}$$

其中, 如果 $f_\eta(y) = 0$, 则设 $f_{\xi|\eta}(x|y) = 0$.

那么,

$$\mathbf{P}\{\xi \in C|\eta = y\} = \int_C f_{\xi|\eta}(x|y)\mathrm{d}x, \quad C \in \mathscr{B}(\mathbb{R}), \tag{19}$$

即 $f_{\xi|\eta}(x|y)$ 是条件概率分布密度.

事实上, 为证明 (19) 式, 只需对于 $B \in \mathscr{B}(\mathbb{R}), A = \{\xi \in C\}$, 证明 (17) 式. 由 §6 的 (43) 式和 (45) 式以及傅比尼定理, 可见

$$\begin{aligned}\int_B \left[\int_C f_{\xi|\eta}(x|y)\mathrm{d}x\right] P_\eta(\mathrm{d}y) &= \int_B \left[\int_C f_{\xi|\eta}(x|y)\mathrm{d}x\right] f_\eta(y)\mathrm{d}y \\ &= \int_{C\times B} f_{\xi|\eta}(x|y) f_\eta(y)\mathrm{d}x\mathrm{d}y = \int_{C\times B} f_{\xi,\eta}(x, y)\mathrm{d}x\mathrm{d}y \\ &= \mathbf{P}\{(\xi, \eta) \in C \times B\} = \mathbf{P}(\{\xi \in C\} \cap \{\eta \in B\}),\end{aligned}$$

于是 (17) 式得证,

如果 $\mathbf{E}\xi$ 存在, 则类似可得

$$\mathbf{E}(\xi|\eta = y) = \int_{-\infty}^{\infty} x f_{\xi|\eta}(x|y)\mathrm{d}x. \tag{20}$$

例 3 假设某设备工作的时间是一非负随机变量 $\eta = \eta(\omega)$, 其分布函数 $F_\eta(y)$ 有密度 $f_\eta(y)$(对于 $y < 0$, 自然 $F_\eta(y) = f_\eta(y) = 0$). 求条件数学期望 $\mathbf{E}(\eta - a|\eta \geqslant a)$, 即在设备已经工作了时间 a 的条件下, 它还能再继续工作的平均时间.

设 $\mathbf{P}\{\eta \geqslant a\} > 0$. 那么, 根据定义 (见第 1 小节) 和 §6 的 (45) 式, 可见

$$\begin{aligned}\mathbf{E}(\eta - a|\eta \geqslant a) &= \frac{\mathbf{E}[(\eta - a)I_{\{\eta\geqslant a\}}]}{\mathbf{P}\{\eta \geqslant a\}} \\ &= \frac{\int_\Omega (\eta - a)I_{\{\eta\geqslant a\}}\mathbf{P}(\mathrm{d}\omega)}{\mathbf{P}\{\eta \geqslant a\}} = \frac{\int_a^\infty (y - a)f_\eta(y)\mathrm{d}y}{\int_a^\infty f_\eta(y)\mathrm{d}y}.\end{aligned}$$

有趣的是, 如果随机变量 η 服从指数分布, 即

$$f_\eta(y) = \begin{cases} \lambda e^{-\lambda y}, & 若\ y \geqslant 0, \\ 0, & 若\ y < 0, \end{cases} \tag{21}$$

则 $\mathbf{E}\eta = \mathbf{E}(\eta|\eta \geqslant 0) = 1/\lambda$, 且对于任意 $a > 0$, $\mathbf{E}(\eta - a|\eta \geqslant a) = 1/\lambda$. 换句话说, 在设备已经工作了时间 a 的条件下, 它还能再继续工作的平均时间, 与 a 的值无关, 就等于 $\mathbf{E}\eta$.

在 (21) 式的条件下, 求条件分布 $\mathbf{P}(\eta - a \leqslant x|\eta \geqslant a)$. 有

$$\begin{aligned} \mathbf{P}(\eta - a \leqslant x|\eta \geqslant a) &= \frac{\mathbf{P}\{a \leqslant \eta \leqslant a + x\}}{\mathbf{P}\{\eta \geqslant a\}} = \frac{F_\eta(a+x) - F_\eta(a) + \mathbf{P}\{\eta = a\}}{1 - F_\eta(a) + \mathbf{P}\{\eta = a\}} \\ &= \frac{[1 - e^{-\lambda(a+x)}] - [1 - e^{-\lambda a}]}{1 - [1 - e^{-\lambda a}]} = \frac{e^{-\lambda a}[1 - e^{-\lambda x}]}{e^{-\lambda a}} = 1 - e^{-\lambda x}. \end{aligned}$$

于是, 条件分布 $\mathbf{P}(\eta - a \leqslant x|\eta \geqslant a)$ 等于无条件分布 $\mathbf{P}\{\eta \leqslant x\}$. 指数分布这一特别好的性质是特有的: 不存在其他分布, 使其密度具有同样性质 $\mathbf{P}(\eta - a \leqslant x|\eta \geqslant a) = \mathbf{P}\{\eta \leqslant x\}$,$a \geqslant 0, 0 \leqslant x < \infty$.

例 4 (蒲丰 [C. de Buffon] 掷针问题) 假设向平面上一无限长和单位宽的 "走廊" (两条直线所夹的区域) 上 (图 29),"随机" 地投掷一单位长度的针. 问针 (至少) 与 "走廊" 的一侧相交的概率如何?

为解决所提 (几何概率的) 问题, 首先说清何谓 "随机" 地投掷一单位长度的针. 以 ξ 表示针落地后其中点到 "走廊" 左侧的距离. 假设在区间 [0,1] 上均匀分布, 而角 θ 在 $[-\pi/2, \pi/2]$ 上均匀 ($P_\theta(\mathrm{d}a) = \mathrm{d}a$) 分布 (图 29). 此外, 假设 ξ 和 θ 独立.

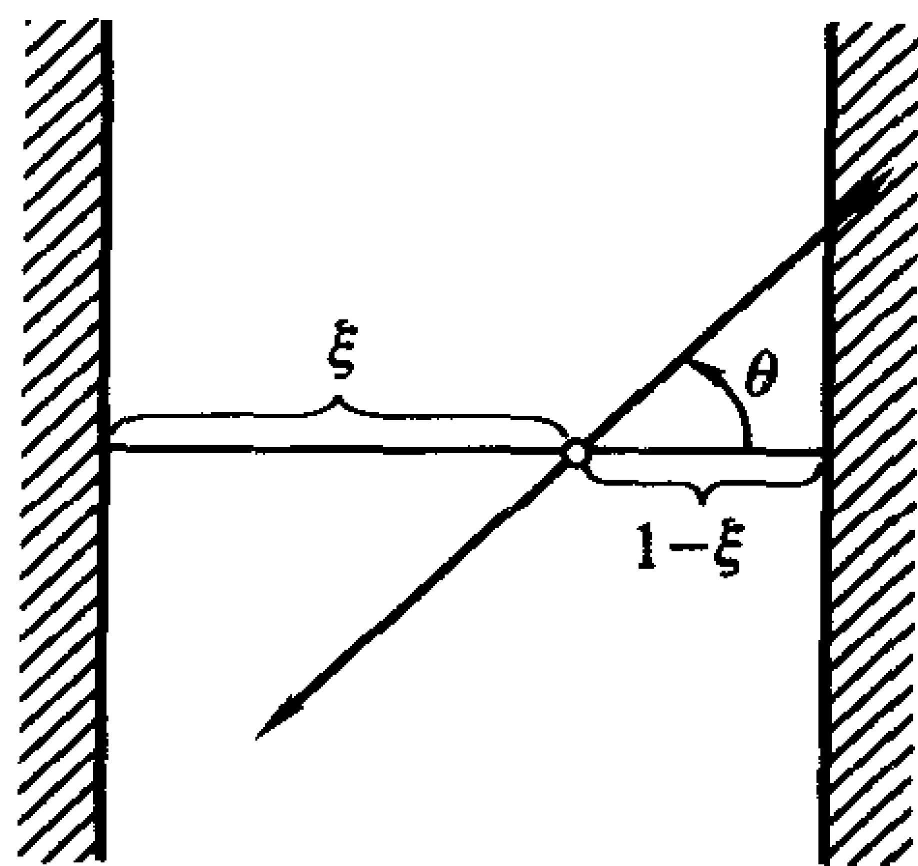


图 29

设 A 是表示事件针与 "走廊" 的某一侧相交. 易见, 如果

$$B = \left\{(a, x) : |a| \leqslant \frac{\pi}{2}, x \in \left[0, \frac{1}{2}\cos a\right] \cup \left[1 - \frac{1}{2}\cos a, 1\right]\right\},$$

则 $A = \{\omega : (\theta, \xi) \in B\}$, 因而所求概率为

$$\mathbf{P}(A) = \mathbf{E}I_A(\omega) = \mathbf{E}I_B(\theta(\omega), \xi(\omega)).$$

由于性质 $\mathbf{G}^*$ 和 (16) 式, 有

$$\begin{aligned}\mathbf{E}I_B(\theta(\omega),\xi(\omega)) &= \mathbf{E}[\mathbf{E}(I_B(\theta(\omega),\xi(\omega))|\theta(\omega))]\\ &= \int_\Omega \mathbf{E}[I_B(\theta(\omega),\xi(\omega))|\theta(\omega)]\mathbf{P}(\mathrm{d}\omega)\\ &= \int_{-\pi/2}^{\pi/2} \mathbf{E}[I_B(\theta(\omega),\xi(\omega))|\theta(\omega)=a]P_\theta(\mathrm{d}a)\\ &= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\mathbf{E}I_B(a,\xi(\omega))\mathrm{d}a = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos\ a\mathrm{d}a = \frac{2}{\pi},\end{aligned}$$

其中用到

$$\mathbf{E}I_B(a,\xi(\omega)) = \mathbf{P}\left\{\xi\in\left[0,\frac{1}{2}\cos\ a\right]\cup\left[1-\frac{1}{2}\cos\ a,1\right]\right\} = \cos\ a.$$

这样, “随机地” 向 “走廊” 中投掷的针落地后, 至少与 “走廊” 的一侧相交的概率等于 $2/\pi$. 这一结果, 可以作为通过试验确定 π 值的基础. 实际上, 假设将针独立地掷 N 次, 并定义随机变量:

$$\eta_i = \begin{cases}1, & \text{若第 } i \text{ 次掷针与走廊侧边相交,}\\ 0, & \text{若第 } i \text{ 次掷针与走廊侧边不交.}\end{cases}$$

那么, 由大数定律 (例如, 见第一章 §5 (6) 式), 可见对于任意 $\varepsilon>0$,

$$\mathbf{P}\left\{\left|\frac{\eta_1+\cdots+\eta_N}{N}-\mathbf{P}(A)\right|>\varepsilon\right\}\to 0,\quad N\to\infty.$$

在此意义上, 频率

$$\frac{\eta_1+\cdots+\eta_N}{N}\approx\mathbf{P}(A)=\frac{2}{\pi},$$

于是

$$\frac{2N}{\eta_1+\cdots+\eta_N}\approx\pi.$$

正是这一公式曾经是用统计方法确定 π 值的基础. 在 1850 年 R. 沃尔弗 (R. Wolf) 掷针 5000 次, 结果得的值 3.1596. 看来, 这种方法, 是在数值分析中利用概率统计方法最早的用法之一, 也就是现代熟知的 “蒙特卡罗法”.

注 所讨论的例 4 (蒲丰掷针问题) 是几何概率的典型问题. 在这类问题中, 相当常见的是, 由诸如 “对称性” 之类的简单几何直观, 可见如何确定 “一组基本事件” 的概率. (对照第一章 §1 第 3 和 4 小节, 以及这一章的 §3.) 下面练习题的第 9~12 题是几何概率的题.

7. 条件概率的正则性 设 $\{\xi_n\}_{n\geqslant1}$ 是非负随机变量序列, 则可见定理 2 的命题 f),

$$\mathbf{E}\left(\sum\xi_n|\mathscr{G}\right)=\sum\mathbf{E}(\xi_n|\mathscr{G})\quad(\text{a.c.}).$$

特别, 若 $B_1, B_2, \cdots$ 是两两不相交的集合, 则

$$\mathbf{P}\left(\sum B_n|\mathscr{G}\right) = \sum \mathbf{P}(B_n|\mathscr{G}) \quad \text{(a.c.)}. \tag{22}$$

重要的是强调这一等式只是几乎必然成立. 因而, 对于固定的 ω 不能将条件概率 $\mathbf{P}(B|\mathscr{G})(\omega)$ 看成 B 的测度. 也许除去一个零测集 $\mathscr{N}$ 之后, 就可以对于每一个 $\omega \in \overline{\mathscr{N}}$, 使得 $\mathbf{P}(\cdot|\mathscr{G})(\omega)$ 是概率测度. 然而, 一般并非如此, 看下面的例子. 对于给定的 $B_1, B_2, \cdots$, 以 $\mathscr{N}(B_1, B_2, \cdots)$ 表示不满足可数可加性 (22) 式的结局 ω 的集合. 那么, 特殊集合 $\mathscr{N}$ 为:

$$\mathscr{N} = \cup \mathscr{N}(B_1, B_2, \cdots), \tag{23}$$

其中对 $\mathscr{F}$ 中一切两两不相交的集合 $B_1, B_2, \cdots$ 求并. 虽然每一个集合的 $\mathbf{P}-$ 测度等于 0, 然而集合 $\mathscr{N}$ 的 $\mathbf{P}-$ 测度可能不等于 0 (因为 (23) 式的并含不可数项). 例如, 个别点的勒贝格测度为 0, 而集合 $\mathscr{N} = [0,1)$ 的测度等于 1, 因为 $\mathscr{N} = [0,1)$ 是单点集 $\{x\}, 0 \leqslant x < 1$, 的不可数的和.

假如对于每一个 $\omega \in \Omega$, 当条件概率 $\mathbf{P}(\cdot|\mathscr{G})(\omega)$ 是概率测度时, 将更为方便. 例如, 因为这时计算条件数学期望 $\mathbf{E}(\xi|\mathscr{G})(\omega)$, 可以利用对测度 $\mathbf{P}(\cdot|\mathscr{G})$ 求平均

$$\mathbf{E}(\xi|\mathscr{G})(\omega) = \int_\Omega \xi(\overline{\omega})\mathbf{P}(\mathrm{d}\overline{\omega}|\mathscr{G})(\omega) \quad \text{(a.c.)}$$

简单地实现 (对照第一章 §8 (10) 式).

现在引进正则条件概率的概念.

定义 6 对于一切 $\omega \in \Omega, B \in \mathscr{F}$, 定义的函数 $P(\omega; B)$, 称做关于 $\sigma-$ 代数 $\mathscr{G} \subseteq \mathscr{F}$ 为正则条件概率, 如果:

a) 对于每一个$\omega \in \Omega, P(\omega; \cdot)$ 是 $\mathscr{F}$ 上的概率测度;

b) 对于每一个 $B \in \mathscr{F}$, $P(\omega; B)$ 作为 ω 的函数是条件概率 $\mathbf{P}(B|\mathscr{G})(\omega)$ 的变式之一, 即

$$P(\omega; B) = \mathbf{P}(\xi|\mathscr{G})(\omega) \quad \text{(a.c.)}.$$

定理 3 设 $P(\omega; B)$ 关于 $\mathscr{G}$ 是正则条件概率, ξ 是可积随机变量, 则

$$\mathbf{E}(\xi|\mathscr{G})(\omega) = \int_\Omega \xi(\widetilde{\omega})P(\omega; \mathrm{d}\widetilde{\omega}) \quad \text{(a.c.)}. \tag{24}$$

证明 如果 $\xi = I_B, B \in \mathscr{F}$, 则要证明的 (24) 式变为

$$\mathbf{P}(\xi|\mathscr{G})(\omega) = P(\omega; B) \quad \text{(a.c.)}.$$

并且由定义 6 的 b), 知此等式成立. 从而, (24) 式对于简单函数成立.

现在设 $\xi \geqslant 0$ 和 $\xi_n \uparrow \xi$, 其中 ξ_n 是简单函数. 那么, 由定理 2 的性质 b), 有

$$\mathbf{E}(\xi|\mathscr{G})(\omega) = \lim_n \mathbf{E}(\xi_n|\mathscr{G})(\omega) \quad \text{(a.c.)}.$$

因为对于每一个 $\omega \in \Omega, P(\omega;\cdot)$ 是测度, 则根据单调收敛定理, 有

$$\lim_n \mathbf{E}(\xi_n|\mathscr{G})(\omega) = \lim_n \int_\Omega \xi_n(\widetilde{\omega})P(\omega;\mathrm{d}\widetilde{\omega}) = \int_\Omega \xi(\widetilde{\omega})P(\omega;\mathrm{d}\widetilde{\omega}).$$

由于 $\xi = \xi^+ - \xi^-$, 故一般情形可以归结为上面已经证明的情形, □

系 设 $\mathscr{G} = \mathscr{G}_\eta$, 其中 η 是随机变量, 而 (ξ,η) 是随机向量, 且具有概率分布密度 $f_{\xi,\eta}(x,y)$. 假设 $\mathbf{E}|g(\xi)| < \infty$, 那么

$$\mathbf{E}[g(\xi)|\eta = y] = \int_{-\infty}^{\infty} g(x) f_{\xi|\eta}(x|y)\mathrm{d}x,$$

其中 $f_{\xi|\eta}(x|y)$ 是条件分布密度 (见 (18) 式).

为表述正则条件概率存在的基本结果, 需要下面的定义.

定义 7 设 $(E,\mathscr{E})$ 是可测空间, $X = X(\omega)$ 是值空间为 E 的随机元, $\mathscr{G}$ 是 $\mathscr{F}$ 的 $\sigma-$ 子代数. 对于 $\omega \in \Omega$ 和 $B \in \mathscr{E}$ 定义的函数 $Q(\omega; B)$, 称做 X 关于 $\sigma-$ 代数 $\mathscr{G}$ 的正则条件分布, 如果:

a) 对于任意 $\omega \in \Omega$, $Q(\omega; B)$ 是 $(E,\mathscr{E})$ 上的概率测度;

b) 对一个每 $B \in \mathscr{E}$, $Q(\omega; B)$ 作为 ω 的函数, 是条件概率 $\mathbf{P}(X \in B|\mathscr{G})(\omega)$ 的变式之一, 即

$$Q(\omega; B) = \mathbf{P}(X \in B|\mathscr{G})(\omega) \quad \text{(a.c.)}.$$

定义 8 设 ξ 是随机变量. 函数 $F(\omega; x), \omega \in \Omega, x \in \mathbb{R}$ 称做 ξ 关于 $\sigma-$ 代数 $\mathscr{G}$ 的正则分布函数, 如果:

a) 对于每一 $\omega \in \Omega, F(\omega; x)$ 是 $\mathbb{R}$ 上的分布函数;

b) 对于每一 $x \in \mathbb{R}, F(\omega; x) = \mathbf{P}(\xi \leqslant x|\mathscr{G})(\omega) \quad \text{(a.c.)}$.

定理 4 随机变量 ξ 关于 $\sigma-$ 代数 $\mathscr{G} \subseteq \mathscr{F}$ 的正则分布函数和正则条件分布永远存在.

证明 对于每一个有理数 $r \in \mathbb{R}$, 设

$$F_r(\omega) = \mathbf{P}(\xi \leqslant r|\mathscr{G})(\omega), \text{ 其中 } \mathbf{P}(\xi \leqslant r|\mathscr{G})(\omega) = \mathbf{E}(I_{\{\xi\leqslant r\}}|\mathscr{G})(\omega)$$

是事件 $\{\xi \leqslant r\}$ 关于 $\mathscr{G}$ 的条件概率的一种变式. 设 $\{r_i\}$ 是 $\mathbb{R}$ 上有理数集. 如果 $r_i < r_j$, 则由于性质 $\mathbf{B}^*$, 有 $\mathbf{P}(\xi \leqslant r_i|\mathscr{G}) \leqslant \mathbf{P}(\xi \leqslant r_j|\mathscr{G})$ (a.c.). 这说明, 若

$$A_{ij} = \{\omega : F_{r_j}(\omega) < F_{r_i}(\omega)\}, \quad A = \bigcup A_{ij},$$

则 $\mathbf{P}(A) = 0$. 即使分布函数 $F_r(\omega), r \in \{r_i\}$, 不单调的集合的测度等于 0.

现在, 设

$$B_i = \{\omega : \lim_{n\to\infty} F_{r_i+1/n}(\omega) \neq F_{r_i}(\omega)\}, \quad B = \bigcup_{i=1}^{\infty} B_i.$$

显然 $I_{\{\xi\leqslant r_i+1/n\}}\downarrow I_{\{\xi\leqslant r_i\}}, n\to\infty$. 因此, 根据定理 2 的命题 a), 有 $F_{r_i+1/n}(\omega)\to F_{r_i}(\omega)$ (a.c.). 说明 (在有理点上) 不右连续点的集合 B 的测度也是 $0:\mathbf{P}(B)=0$.

其次, 设

$$C=\{\omega:\lim_{n\to\infty}F_n(\omega)\neq 1\}\cup\{\omega:\lim_{n\to-\infty}F_n(\omega)\neq 0\}.$$

那么, 由于 $\{\xi\leqslant n\}\uparrow\Omega, n\to\infty$, 而 $\{\xi\leqslant n\}\downarrow\varnothing, n\to-\infty$, 则 $\mathbf{P}(C)=0$.

现在, 设

$$F(\omega;x)=\begin{cases}\lim\limits_{r\downarrow x}F_r(\omega), & 若\ \omega\notin A\cup B\cup C,\\ G(x), & 若\ \omega\in A\cup B\cup C,\end{cases}$$

其中 $G(x)$ 是 $\mathbb{R}$ 上的任意分布函数. 下面证明函数 $F(\omega;x)$ 满足定义 8.

设 $\omega\notin A\bigcup B\bigcup C$, 那么 $F(\omega;x)$ 显然是 x 的不减函数. 对于 $x<x'\leqslant r$, 当 $r\downarrow x$ 时, 则

$$F(\omega;x)\leqslant F(\omega;x')\leqslant F(\omega;r)=F_r(\omega)\downarrow F(\omega;x).$$

因此 $F(\omega;x)$ 右连续. 类似, 有

$$\lim_{x\to\infty}F(\omega;x)=1,\ \lim_{x\to-\infty}F(\omega;x)=0.$$

由于当 $\omega\in A\bigcup B\bigcup C$ 时, $F(\omega;x)=G(x)$, 则对于每个 $\omega\in\Omega, F(\omega;x)$ 是 $\mathbb{R}$ 上的分布函数, 即满足定义 6 中的条件 a).

根据构造, $\mathbf{P}(\xi\leqslant r|\mathscr{G})(\omega)=F_r(\omega)=F(\omega;r)$. 如果 $r\downarrow x$, 则由于已经证明的 $F(\omega;x)$ 的右连续性, 可见对于一切 $\omega\in\Omega$, 有 $F(\omega;r)\downarrow F(\omega;x)$; 由定理 2 的命题 a) 知, 几乎处处有 $\mathbf{P}(\xi\leqslant r|\mathscr{G})(\omega)\to\mathbf{P}(\xi\leqslant x|\mathscr{G})(\omega)$. 因此, $F(\omega;x)=\mathbf{P}(\xi\leqslant x|\mathscr{G})(\omega)$ (a.c.). 于是, 定义 8 的性质 b) 得证.

现在证明变量 ξ 关于 $\mathscr{G}$ 的正则条件分布存在.

假设 $F(\omega;r)$ 是上面建立的函数. 设

$$Q(\omega;B)=\int_B F(\omega;\mathrm{d}x),$$

其中的积分是勒贝格 - 斯蒂尔切斯积分. 由勒贝格 - 斯蒂尔切斯积分性质可见 (见 §6 第 8 小节), 对于每个固定的 $\omega\in\Omega, Q(\omega;B)$ 是 B 的测度. 我们利用 "**适当集合原理**" (见 §2 第 1 小节), 证明 $Q(\omega;B)$ 是条件概率 $\mathbf{P}(B|\mathscr{G})(\omega)$ 的变式.

设 $\mathscr{C}$ 是 $\mathscr{B}(\mathbb{R})$ 中使 $Q(\omega;B)=\mathbf{P}(\xi\in B|\mathscr{G})(\omega)$ (a.c.) 集合 B 的全体. 由于 $F(\omega;x)$ 几乎必然等于 $\mathbf{P}(\xi\leqslant x|\mathscr{G})(\omega)$, 可见形如 $B=(-\infty,x], x\in\mathbb{R}$, 的集合 B 属于集系 $\mathscr{C}$. 说明集系 $\mathscr{C}$ 也包括一切形如 $(a,b]$ 的区间, 以及由有限个不相交形如 $(a,b]$ 的区间的和组成的代数 $\mathscr{A}$. 那么, 由测度 $Q(\omega;B)$ (其中 ω 固定) 的连续性, 以及定理 2 的命题 b), 可见 $\mathscr{C}$ 是单调类. 因为 $\mathscr{A}\subseteq\mathscr{C}\subseteq\mathscr{B}(\mathbb{R})$, 所以由 §2 的定理 1, 有

$$\mathscr{B}(\mathbb{R})=\sigma(\mathscr{A})\subseteq\sigma(\mathscr{C})=\mu(\mathscr{C})=\mathscr{C}\subseteq\mathscr{B}(\mathbb{R}).$$

于是 $\mathscr{C}=\mathscr{B}(\mathbb{R})$. □

利用不复杂的拓扑分析, 可以将定理 4 关于存在正则条件分布的论断推广到在博雷尔空间取值的随机元. 现在引进如下定义.

定义 9 可测空间 $(E,\mathscr{E})$ 称做博雷尔空间, 如果它按博雷尔等价于数轴的某一博雷尔子集, 即存在一一映射 $\varphi=\varphi(e)$: $(E,\mathscr{E})\to(\mathbb{R},\mathscr{B}(\mathbb{R}))$, 使得

1) $\varphi(E)\equiv\{\varphi(e):e\in\mathscr{E}\}$ 是 $\mathscr{B}(\mathbb{R})$ 的某一集合;

2) φ 为 $\mathscr{E}-$ 可测: $[\varphi^{-1}(A)\in\mathscr{E},A\in\varphi(E)\bigcap\mathscr{B}(\mathbb{R})]$;

3) φ^{-1} 为 $\mathscr{B}(\mathbb{R})/\mathscr{E}-$ 可测: $\{\varphi(B)\in\varphi(E)\bigcap\mathscr{B}(\mathbb{R}),B\in\mathscr{E}\}$.

定理 5 *设 $X=X(\omega)$ 是在博雷尔空间 $(E,\mathscr{E})$ 取值的随机元. 那么, 存在 X 的关于 $\sigma-$ 代数 $\mathscr{G}\subseteq\mathscr{F}$ 的正则条件分布.*

证明 设 $\varphi=\varphi(e)$ 是定义 9 中的函数. 由于定义 9 中的 2), $\varphi(X(\omega))$ 是随机变量. 故由定理 4, 随机变量 $\varphi(X(\omega))$ 关于 $\mathscr{G}$ 的条件分布 $Q(\omega;A),A\in\varphi(E)\bigcap\mathscr{B}(\mathbb{R})$ 有定义.

引进函数 $\widetilde{Q}(\omega;B)=Q(\omega;\varphi(B)),B\in\mathscr{E}$. 由定义 9 的 3) $\varphi(B)\in\varphi(E)\bigcap\mathscr{B}(\mathbb{R})$, 因而 $\widetilde{Q}(\omega;B)$ 有定义. 显然, 对于每一个 $\omega,\widetilde{Q}(\omega;B)$ 关于 $B(B\in\mathscr{E})$ 是测度. 现在固定 $B\in\mathscr{E}$. 由于 $\varphi=\varphi(e)$ 是一一映射, 有

$$\widetilde{Q}(\omega;B)=Q(\omega;\varphi(B))=\mathbf{P}\{\varphi(X)\in\varphi(B)|\mathscr{G}\}(\omega)=\mathbf{P}\{X\in B|\mathscr{G}\}(\omega)\quad\text{(a.c.)}.$$

于是, $\widetilde{Q}(\omega;B)$ 是 X 的关于 $\mathscr{G}$ 的条件正则分布. □

系 *设 $X=X(\omega)$ 是在完全可分度量空间 $(E,\mathscr{E})$ 取值的随机元. 那么, 存在 X 的关于 $\mathscr{G}$ 的正则条件分布. 特别, 对于空间 $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 和 $(\mathbb{R}^\infty,\mathscr{B}(\mathbb{R}^\infty))$ 存在这样的分布.*

其证明可以从两方面得到: 第一, 定理 5; 第二, 由拓扑中著名结果知, 这样的空间是博雷尔空间.

8. 贝叶斯定理的推广 由上面阐述的条件数学期望的理论, 可以得到在统计学中用到的贝叶斯定理的推广.

回忆. $\mathscr{D}=\{A_1,\cdots,A_n\}(\mathbf{P}(A_i)>0)$ 是空间 Ω 的分割, 则由第一章 §3 (9) 式的贝叶斯定理知, 对于任意 $B,\mathbf{P}(B)>0$, 有

$$\mathbf{P}(A_i|B)=\frac{\mathbf{P}(A_i)\mathbf{P}(B|A_i)}{\sum\limits_{j=1}^{n}\mathbf{P}(A_j)\mathbf{P}(B|A_j)}.\tag{25}$$

因此, 如果 $\theta=\sum_{i=1}^{n}a_iI_{A_i}$ 是离散型随机变量, 则根据第一章 §8 的 (10) 式, 有

$$\mathbf{E}[g(\theta)|B]=\frac{\sum\limits_{i=1}^{n}g(a_i)\mathbf{P}(A_i)\mathbf{P}(B|A_i)}{\sum\limits_{j=1}^{n}\mathbf{P}(A_j)\mathbf{P}(B|A_j)},\tag{26}$$

或

$$\mathbf{E}[g(\theta)|B] = \frac{\int_{-\infty}^{\infty} g(a)\mathbf{P}(B|\theta = a)P_\theta(\mathrm{d}a)}{\int_{-\infty}^{\infty} \mathbf{P}(B|\theta = a)P_\theta(\mathrm{d}a)}. \tag{27}$$

其中 $P_\theta(A) = \mathbf{P}\{\theta \in A\}$.

由刚引进的 $\mathbf{E}[g(\theta)|B]$ 的定义, 不难证明, (27) 式对于任何事件 $B, \mathbf{P}(B) > 0$, 以及随机变量 $\theta = \theta(\omega)$ 和函数 $g = g(a)$ 且 $\mathbf{E}|g(\theta)| < \infty$, 仍然成立.

现在, 对于某 $\sigma-$ 代数 $\mathscr{G}(\mathscr{G} \subseteq \mathscr{F})$ 的条件数学期望 $\mathbf{E}[g(\theta)|\mathscr{G}]$, 考虑 (27) 式的类似.

设

$$\mathbf{Q}(B) = \int_B g(\theta(\omega))\mathbf{P}(\mathrm{d}\omega), \quad B \in \mathscr{G}, \tag{28}$$

则由 (4) 式可见

$$\mathbf{E}[g(\theta)|\mathscr{G}] = \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}(\omega). \tag{29}$$

与 $\sigma-$ 代数 $\mathscr{G}$ 同时, 再考虑 $\sigma-$ 代数 $\mathscr{G}_\theta$, 则由 (5) 式可见

$$\mathbf{P}(B) = \int_\Omega \mathbf{P}(B|\mathscr{G}_\theta)\mathrm{d}\mathbf{P}, \tag{30}$$

或根据在勒贝格积分号下的变量替换公式, 有

$$\mathbf{P}(B) = \int_{-\infty}^{\infty} \mathbf{P}(B|\theta = a)P_\theta(\mathrm{d}a). \tag{31}$$

由于

$$\mathbf{Q}(B) = \mathbf{E}[g(\theta)I_B] = \mathbf{E}[g(\theta)\mathbf{E}(I_B|\mathscr{G}_\theta)],$$

可见

$$\mathbf{Q}(B) = \int_{-\infty}^{\infty} g(a)\mathbf{P}(B|\theta = a)P_\theta(\mathrm{d}a). \tag{32}$$

现在假设这些条件概率 $\mathbf{P}(B|\theta = a)$ 都是正则的, 并且可以表示为

$$\mathbf{P}(B|\theta = a) = \int_B \rho(\omega, a)\lambda(\mathrm{d}\omega), \tag{33}$$

其中 $\rho = \rho(\omega; a)$ 是非负 (对两个变量) 可测的函数, 而 λ 是 $(\Omega, \mathscr{G})$ 上的某 $\sigma-$ 有限测度. 设 $\mathbf{E}|g(\theta)| < \infty$. 现在证明, 依概率 1, 有 (*广义贝叶斯定理*)

$$\mathbf{E}[g(\theta)|\mathscr{G}](\omega) = \frac{\int_{-\infty}^{\infty} g(a)\rho(\omega; a)P_\theta(\mathrm{d}a)}{\int_{-\infty}^{\infty} \rho(\omega; a)P_\theta(\mathrm{d}a)}. \tag{34}$$

为证明 (34) 式, 需要下面的引理.

引理 设 $(\Omega,\mathscr{F})$ 是某个可测空间.

a) 设 μ 和 $\lambda(\mu\ll\lambda)$ 是 $\sigma-$ 有限测度, $f=f(\omega)$ 是 $\mathscr{F}-$ 可测函数, 则

$$\int_{\Omega} f\mathrm{d}\mu=\int_{\Omega} f\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\mathrm{d}\lambda \tag{35}$$

(即如果其中一个积分存在, 则另一个也存在, 并且二者相等).

b) 如果 ν 是带符号的测度, 而 μ 和 $\lambda(\mu\ll\lambda)$ 是 $\sigma-$ 有限测度, 且 $\nu\ll\mu\ll\lambda$, 则

$$\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}=\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\times\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\quad(\lambda-\text{a.c.}) \tag{36}$$

和

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}=\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\bigg/\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\quad(\mu-\text{a.c.}). \tag{37}$$

证明 a) 由于

$$\mu(A)=\int_A\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda,\ A\in\mathscr{F},$$

可见对于任意简单函数 $f=\sum f_iI_{A_i}$, (35) 式自然成立. 一般情形, 由表示 $f=f^+-f^-$ 和单调收敛定理可以得到 (35) 式 (对照 §6 (39) 式的证明).

b) 由命题 a), 其中设 $f=\mathrm{d}\nu/\mathrm{d}\mu$, 可得

$$\nu(A)=\int_A\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)\mathrm{d}\mu=\int_A\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda.$$

因此 $\nu\ll\lambda$, 故

$$\nu(A)=\int_A\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\mathrm{d}\lambda.$$

由于集合 A 的任意性和 §6 的性质 **I**, 可得 (36) 式.

由 (36) 式, 以及如下等式可得 (37) 式:

$$\mu\left\{\omega:\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}=0\right\}=\int_{\{\omega:\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}=0\}}\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda=0$$

(在集合 $\{\omega:\mathrm{d}\mu/\mathrm{d}\lambda=0\}$ 上 (37) 式的右侧可以任意决定, 例如, 令其等于 0). □

为证明 (34) 式注意到, 由傅比尼定理和假设 (33), 有

$$\mathbf{Q}(B)=\int_B\left[\int_{-\infty}^{\infty}g(a)\rho(\omega;a)P_\theta(\mathrm{d}a)\right]\lambda(\mathrm{d}\omega), \tag{38}$$

$$\mathbf{P}(B)=\int_B\left[\int_{-\infty}^{\infty}\rho(\omega;a)P_\theta(\mathrm{d}a)\right]\lambda(\mathrm{d}\omega). \tag{39}$$

那么, 由引理, 有

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}=\frac{\mathrm{d}\mathbf{Q}/\mathrm{d}\lambda}{\mathrm{d}\mathbf{P}/\mathrm{d}\lambda}\quad(\mathbf{P}-\text{a.c.}),$$

由此并考虑到 (38),(39) 和 (29) 式, 得 (34) 式.

注 如果代替随机变量 θ, 考虑在某一可测空间 $(E,\mathscr{E})$ 取值的随机元, 则 (34) 式仍然成立 (只要把在 $\mathbb{R}$ 上积分换成在 E 上积分).

下面讨论 (34) 式的某些特殊情形.

假设 $\sigma-$ 代数 $\mathscr{G}$ 由随机变量 ξ 生成: $\mathscr{G}=\sigma(\xi)$.

假设

$$\mathbf{P}(\xi\in A|\theta=a)=\int_A q(x;a)\lambda(\mathrm{d}x),\quad A\in\mathscr{B}(\mathbb{R}),\tag{40}$$

其中 $q=q(x;a)$ 是关于两个变量非负可测的函数, λ 是 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的某一 $\sigma-$ 有限测度. 那么, 由勒贝格积分在符号下的变量替换和 (34) 式, 可得

$$\mathbf{E}[g(\theta)|\xi=x]=\frac{\displaystyle\int_{-\infty}^{\infty}g(a)q(x;a)P_\theta(\mathrm{d}a)}{\displaystyle\int_{-\infty}^{\infty}q(x;a)P_\theta(\mathrm{d}a)}.\tag{41}$$

特别, 设 $(\theta,\xi),\theta=\sum a_iI_{A_i},\xi=\sum x_jI_{B_j}$, 是两个离散型随机变量. 那么, 作为 λ 选计数测度 $(\lambda(\{x_i\})=1,i=1,2,\cdots)$, 由 (40) 式, 得 (对照 (26) 式)

$$\mathbf{E}[g(\theta)|\xi=x_j]=\frac{\displaystyle\sum_i g(a_i)\mathbf{P}\{\xi=x_j|\theta=a_i\}\mathbf{P}\{\theta=a_i\}}{\displaystyle\sum_i\mathbf{P}\{\xi=x_j|\theta=a_i\}\mathbf{P}\{\theta=a_i\}}.\tag{42}$$

现在设 (θ,ξ) 是密度为 $f_{\theta,\xi}(a,x)$ 的两个绝对连续随机变量, 则由于 (19) 式, 可见 (40) 式成立, 其中 $g(x,a)=f_{\xi|\theta}(x|a)$, 而 λ 是勒贝格测度. 因此

$$\mathbf{E}[g(\theta)|\xi=x]=\frac{\displaystyle\int_{-\infty}^{\infty}g(a)f_{\xi|\theta}(x|a)f_\theta(a)\mathrm{d}a}{\displaystyle\int_{-\infty}^{\infty}f_{\xi|\theta}(x|a)f_\theta(a)\mathrm{d}a}.\tag{43}$$

9. 条件数学期望的换算公式 我们再引进广义贝叶斯定理 (见 (34) 式) 的一种形式, 其 (下面的) 提法在与概率测度变换有关的问题中特别适宜.

定理 6 假设 $\mathbf{P}$ 和 $\widetilde{\mathbf{P}}$ 是可测空间 $(\Omega,\mathscr{F})$ 上的两个概率测度, 且测度 $\widetilde{\mathbf{P}}$ 关于测度 $\mathbf{P}$ 绝对连续 (记作 $\widetilde{\mathbf{P}}\ll\mathbf{P}$), 而 $\mathrm{d}\widetilde{\mathbf{P}}/\mathrm{d}\mathbf{P}$ 是测度 $\widetilde{\mathbf{P}}$ 关于测度 $\mathbf{P}$ 的拉东 – 尼科迪姆导数. 设 $\mathscr{G}$ 是 $\mathscr{F}(\mathscr{G}\subseteq\mathscr{F})$ 的子 $\sigma-$ 代数; $\mathbf{E}(\cdot|\mathscr{G})$ 和 $\widetilde{\mathbf{E}}(\cdot|\mathscr{G})$ 相应为按测度 $\mathbf{P}$ 和测度 $\widetilde{\mathbf{P}}$ 关于 $\mathscr{G}$ 的条件数学期望; ξ 是非负 ($\mathscr{F}-$ 可测的) 随机变量, 则下列 "条件数学期望的换算公式" 成立:

$$\widetilde{\mathbf{E}}(\xi|\mathscr{G})=\frac{\mathbf{E}\left(\xi\,\dfrac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\bigg|\,\mathscr{G}\right)}{\mathbf{E}\left(\dfrac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\bigg|\,\mathscr{G}\right)}(\widetilde{\mathbf{P}}-\text{a.c.}).\tag{44}$$

对于使条件数学期望 $\widetilde{\mathbf{E}}(\xi|\mathscr{G})$ 存在的一切随机变量 ξ, 公式 (44) 成立.

证明 首先注意到, 事件

$$\left\{\omega : \mathbf{E}\left(\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)=0\right\}$$

的 $\widetilde{\mathbf{P}}$ 测度 (以及 $\mathbf{P}$ 测度) 等于 0. 事实上, 如果 $A\in\mathscr{G}$, 则

$$\int_A \mathbf{E}\left(\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)\mathrm{d}\mathbf{P}=\int_A \frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\mathrm{d}\mathbf{P}=\int_A \mathrm{d}\widetilde{\mathbf{P}}=\widetilde{\mathbf{P}}(A).$$

因而

$$\widetilde{\mathbf{P}}(A)=\widetilde{\mathbf{P}}\left\{\omega : \mathbf{E}\left(\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)=0\right\}=0.$$

设 $\xi\geqslant 0$. 根据条件数学期望的定义, $\widetilde{\mathbf{E}}(\xi|\mathscr{G})$ 是 $\mathscr{G}-$ 可测随机变量, 且对于任意 $A\in\mathscr{G}$, 满足

$$\widetilde{\mathbf{E}}[I_A\widetilde{\mathbf{E}}(\xi|\mathscr{G})]=\widetilde{\mathbf{E}}[I_A\xi)]. \tag{45}$$

因此, 为证明 (44) 式, 只需证明 (44) 式右侧的 ($\mathscr{G}-$ 可测) 随机变量满足下面的不等式:

$$\widetilde{\mathbf{E}}\left[I_A\times\frac{\mathbf{E}\left(\xi\,\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)}{\mathbf{E}\left(\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)}\right]=\widetilde{\mathbf{E}}[I_A\xi]. \tag{46}$$

利用条件数学期望的性质, 以及 §6 的 (39) 式, 可得

$$\begin{aligned}
&\widetilde{\mathbf{E}}\left[I_A\times\frac{\mathbf{E}\left(\xi\,\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)}{\mathbf{E}\left(\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)}\right]=\widetilde{\mathbf{E}}\left[I_A\times\frac{\mathbf{E}\left(\xi\,\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)}{\mathbf{E}\left(\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)}\times\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\right]\\
&=\widetilde{\mathbf{E}}\left[I_A\times\frac{\mathbf{E}\left(\xi\,\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)}{\mathbf{E}\left(\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)}\times\mathbf{E}\left(\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)\right]=\widetilde{\mathbf{E}}\left[I_A\times\mathbf{E}\left(\xi\,\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{G}\right)\right]\\
&=\widetilde{\mathbf{E}}\left[I_A\xi\frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{P}}\right]=\widetilde{\mathbf{E}}[I_A\xi],
\end{aligned}$$

从而, 对于非负随机变量 ξ, 证明了 (45) 式. 对于任意可积随机变量 ξ, 一般情形的证明, 完全类似于 §6 的 (39) 式的证明. □

10. 充分统计量和因子分解定理　上面引进的广义贝叶斯定理 ((34),(41) 和 (43) 式), 是数理统计中 "贝叶斯方法" 的基本工具之一. 它回答如下问题: 根据对与随机参数 θ 相联系的随机变量 ξ 的观测结果, 如何重新分配关于随机变量 θ 的知识.

下面将讨论条件数学期望概念在由观测结果估计未知参数问题中的另外一种应用. (需要强调: 与上面讨论的 θ 是随机变量的情形不同, 现在 θ 是事先给定的参数集 Θ 中的某一普通参数 (对照第一章 §7)).

实质上, 这里说的是关于数理统计中的一个重要概念 —— *充分子 $\sigma-$ 代数* (或 *$\sigma-$ 子代数*) 的概念.

设 $(\Omega, \mathscr{F})$ 是某一可测空间, $\mathscr{P} = \{\mathbf{P}_\theta, \theta \in \Theta\}$ 是依赖于参数的概率测度 $\mathbf{P}_\theta$ 族, 其中 θ 属于*参数的集合* Θ. 常称组合 $\mathscr{E} = (\Omega, \mathscr{F}, \mathscr{P})$ 为*概率 - 统计模型*或*概率 - 统计试验* (对照第一章 §7).

为说明下面引进的定义 10, 假设有结局 ω 的某一 $\mathscr{F}-$ 可测函数 (统计量), 和由该函数 $T = T(\omega)$ 诱导的 $\sigma-$ 代数 $\mathscr{G} = \sigma(T(\omega))$. 显然, $\mathscr{G} \subseteq \mathscr{F}$, 且一般在 $\mathscr{F}$ 中可能有不属于 $\mathscr{G}$ 的元素 (即 $\mathscr{F}$ 比 $\mathscr{G}$ "丰富"). 但是, 完全可能出现这样的情形: 从定义的角度对什么样的参数 θ "作用" 使得除了关于 $T = T(\omega)$ 的知识之外, 任何关于参数 θ 的其他信息, 没有也不需要有. 在这种意义上, 统计量 T 自然应称为*充分的*.

注 1　在 $T(\omega) = \omega$ 的情形下, 即已知的是结局本身 (而不是其函数), 可以指出如下两种极端的情形.

一种是对于一切 $\theta \in \Theta$, 所有概率 $\mathbf{P}_\theta$ *都相等*. 显然, 在这种情形下, 任何结局 ω 都不能提供关于参数 θ 的值的任何信息.

另一种情形是, 所有概率 $\mathbf{P}_\theta$ 的*承载子*, "处于" $\mathscr{F}$ 的不同子集中 (即, 对于任何两个参数 θ_1 和 θ_2, 测度 $\mathbf{P}_{\theta_1}$ 和 $\mathbf{P}_{\theta_2}$ 是*奇异的*: 存在两个集合 (承载子) Ω_{θ_1} 和 Ω_{θ_2}, 使

$$\mathbf{P}_{\theta_1}(\Omega \backslash \Omega_{\theta_1}) = 0, \quad \mathbf{P}_{\theta_2}(\Omega \backslash \Omega_{\theta_2}) = 0, \quad \Omega_{\theta_1} \cap \Omega_{\theta_2} = \varnothing$$

成立). 在这种情形下, 参数 θ 的值由所得结局 ω "唯一" 确定.

对于这两种极端情形没什么兴趣. 人们感兴趣的是, 所有测度 $\mathbf{P}_\theta$ *都等价*的情形 (而这时其承载子不可区分).

现在, 我们把 $\sigma-$ 代数 $\mathscr{G} = \sigma(T(\omega))$, 以及更一般地把*任何*子 $\sigma-$ 代数 $\mathscr{G} \subseteq \mathscr{F}$, 视为在试验结局为 $\omega \in \Omega$ 时, 所获得的 "信息".

有了这样的信息 $\mathscr{G}$, (像贝叶斯定理的情形一样) 自然关心概率 $\mathbf{P}_\theta$ 如何变化, 即相应的*条件概率* $\mathbf{P}_\theta(\cdot|\mathscr{G})(\omega)$ 如何. 这时 (与注 1 的第一种情形类似) 明显, 假如这些概率对于所有 ω 不依赖于 θ, 与由 "信息" $\mathscr{G}$ 可以得到的相比, 则由 "更多的信息" $\widetilde{\mathscr{G}} \supseteq \mathscr{G}$ 得不到关于参数 θ 的任何更新的内容.

这时, "信息" $\mathscr{G}$ 可以称为*详尽无遗的*, 或者习惯上, 称做*充分的*.

下面的定义表达了上面的论断.

定义 10 设 $E=(\Omega,\mathscr{F},\mathscr{P})$ 为概率 - 统计模型, $\mathscr{P}=\{\mathbf{P}_\theta,\theta\in\Theta\}$ 和 $\mathscr{G}$ 是 $\mathscr{F}(\mathscr{G}\subseteq\mathscr{F})$的子 $\sigma-$ 代数. 称子 $\sigma-$ 代数 $\mathscr{G}$ 对于 $\mathscr{P}$ 族是*充分的*, 如果存在不依赖于条件概率 $\mathbf{P}_\theta(\cdot|\mathscr{G})(\omega)(\theta\in\Theta,\omega\in\Omega)$ 的变式, 即存在函数 $P(A;\omega),A\in\mathscr{F},\omega\in\Omega$, 使对于一切 $A\in\mathscr{F}$ 和 $\theta\in\Theta$, 有

$$\mathbf{P}_\theta(A|\mathscr{G})(\omega)=P(A;\omega),(\mathbf{P}_\theta-\text{a.c.}),\tag{47}$$

即对于一切 $\theta\in\Theta$, $P(A;\omega)$ 是 $\mathbf{P}_\theta(A|\mathscr{G})(\omega)$ 的变式.

如果 $\mathscr{G}=\sigma(T(\omega))$, 则统计量 $T=T(\omega)$ 称做 $\mathscr{P}$ 族的*充分统计量*.

注 2 在统计研究中, 求*充分统计量*的重要性, 在于力图寻找结局 ω 的这样函数 $T=T(\omega)$, 使之在*保留* (关于参数 θ 值的) *信息*的情况下*压缩数据*. 例如, 设想对于很大的 n, 有 $\omega=(x_1,x_2,\cdots,x_n),x_i\in\mathbb{R}$. 那么, 由于现有数据 $x_1,x_2,\cdots,x_n$ 的维数 n 很大, 使建立参数 θ 的 "好" 估计量 (例如, 第一章 §7) 可能成为很困难的问题. 不过, 完全可能出现这样的情形 (这在第一章 §7 我们已经见到), 为建立参数的 "好" 估计量, 完全不需要知道*原始数据* $x_1,x_2,\cdots,x_n$, 只需构造统计量的数据的*和*: $T(\omega)=x_1+x_2+\cdots+x_n$.

显然, 这样的统计量确实可以本质地压缩了数据 (与计算量), 同时对于建立参数 θ 的 "好" 估计量, 也是充分的.

下面的*因子分解定理*提供了, 保障 $\sigma-$ 代数 $\mathscr{G}$ 对于概率测度 $\mathbf{P}_\theta$ 族 $\mathscr{P}=\{\mathbf{P}_\theta,\theta\in\Theta\}$ 的充分性的条件.

定理 7 *设 $\mathscr{P}=\{\mathbf{P}_\theta,\theta\in\Theta\}$ 是被控制测度族, 即存在 $(\Omega,\mathscr{F})$ 上的 $\sigma-$ 有限测度 λ, 使对于 $\theta\in\Theta$, 测度 $\mathbf{P}_\theta$ 关于 λ 绝对连续 ($\mathbf{P}_\theta\ll\lambda$). 设*

$$g_\theta^{(\lambda)}(\omega)=\frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\lambda}(\omega)$$

是测度 $\mathbf{P}_\theta$ 关于测度 λ 的拉东 – 尼科迪姆导数.

子 $\sigma-$ 代数 $\mathscr{G}$ 对于概率测度族 $\mathscr{P}$ 是充分的, 当且仅当函数 $g_\theta^{(\lambda)}(\omega)$ 有因子分解: 存在非负函数 $\widehat{g}_\theta^{(\lambda)}(\omega)$ 和 $h(\omega)$, 其中 $g_\theta^{(\lambda)}(\omega)$ 是 $\mathscr{G}-$ 可测函数, $h(\omega)$ 是 $\mathscr{F}-$ 可测函数, 且对于一切 $\theta\in\Theta$, 有

$$g_\theta^{(\lambda)}(\omega)=\widehat{g}_\theta^{(\lambda)}(\omega)h(\omega)\quad(\lambda-\text{a.c.}).\tag{48}$$

假如把测度 λ 取为测度 $\mathbf{P}_{\theta_0}$, 其中 θ_0 是 Θ 的某个参数, 则对于 $\sigma-$ 代数 $\mathscr{G}$ 的充分性, 充分和必要条件是: 导数

$$g_\theta^{(\theta_0)}(\omega)=\frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mathbf{P}_{\theta_0}}(\omega)$$

本身为 $\mathscr{G}-$ 可测的.

证明　1) 充分性. 根据假设作为控制测度 λ 是 $\sigma-$ 有限测度. 表示存在 $\mathscr{F}-$ 可测不相交集合 $\Omega_k, k \geqslant 1$, 使 $\Omega = \sum_{k \geqslant 1} \Omega_k$ 且 $0 < \lambda(\Omega_k) < \infty, k \geqslant 1$.

建立测度

$$\widetilde{\lambda}(\cdot) = \sum_{k \geqslant 1} \frac{1}{2^k} \frac{\lambda(\Omega_k \cap \cdot)}{1 + \lambda(\Omega_k)}.$$

该测度有限, $\widetilde{\lambda}(\Omega) < \infty$ 且 $\widetilde{\lambda}(\Omega) > 0$. 不失普遍性, 可以认为 λ 是概率测度: $\widetilde{\lambda}(\Omega) = 1$.

那么, 由条件数学期望的换算公式 (44), 对于任意 $\mathscr{F}-$ 可测有限随机变量 $X = X(\omega)$, 有

$$\mathbf{E}_\theta(X|\mathscr{G}) = \frac{\mathbf{E}_{\widetilde{\lambda}}\left(X \left.\frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\widetilde{\lambda}}\right| \mathscr{G}\right)}{\mathbf{E}_{\widetilde{\lambda}}\left(\left.\frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\widetilde{\lambda}}\right| \mathscr{G}\right)} \quad (\mathbf{P}_\theta - \text{a.c.}). \tag{49}$$

根据 (48) 式, 有

$$g_\theta^{(\widetilde{\lambda})} = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\widetilde{\lambda}} = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\lambda} \times \frac{\mathrm{d}\lambda}{\mathrm{d}\widetilde{\lambda}} = g_\theta^{(\lambda)} \frac{\mathrm{d}\lambda}{\mathrm{d}\widetilde{\lambda}} = \widehat{g}_\theta^{(\lambda)} h \frac{\mathrm{d}\lambda}{\mathrm{d}\widetilde{\lambda}}. \tag{50}$$

因此 (49) 式有如下形式:

$$\mathbf{E}_\theta(X|\mathscr{G}) = \frac{\mathbf{E}_{\widetilde{\lambda}}\left(X \widehat{g}_\theta^{(\lambda)} h \left.\frac{\mathrm{d}\lambda}{\mathrm{d}\widetilde{\lambda}}\right| \mathscr{G}\right)}{\mathbf{E}_{\widetilde{\lambda}}\left(\widehat{g}_\theta^{(\lambda)} h \left.\frac{\mathrm{d}\lambda}{\mathrm{d}\widetilde{\lambda}}\right| \mathscr{G}\right)} \quad (\mathbf{P}_\theta - \text{a.c.}). \tag{51}$$

但是 $\widehat{g}_\theta^{(\lambda)}$ 为 $\mathscr{G}-$ 可测, 且

$$\mathbf{E}_\theta(X|\mathscr{G}) = \frac{\mathbf{E}_{\widetilde{\lambda}}\left(X\ h \left.\frac{\mathrm{d}\lambda}{\mathrm{d}\widetilde{\lambda}}\right| \mathscr{G}\right)}{\mathbf{E}_{\widetilde{\lambda}}\left(h \left.\frac{\mathrm{d}\lambda}{\mathrm{d}\widetilde{\lambda}}\right| \mathscr{G}\right)} \quad (\mathbf{P}_\theta - \text{a.c.}). \tag{52}$$

这里的右侧与 θ 无关, 从而性质 (47) 成立. 于是, 根据定义 10, $\sigma-$ 代数 G 是充分的.

2) 必要性. 为证明必要性需要有补充假设: 测度族 $\mathscr{P} = \{\mathbf{P}_\theta, \theta \in \Theta\}$ 不但被某 $\sigma-$ 有限测度 λ 控制, 而且存在某个 $\theta_0 \in \Theta$, 使一切测度 $\mathbf{P}_\theta \ll \mathbf{P}_{\theta_0}$, 即对于任意 $\theta \in \Theta$, 测度 $\mathbf{P}_\theta$ 关于测度 $\mathbf{P}_{\theta_0}$ 绝对连续. (一般情形的证明变得比较复杂, 参见 [106] 的定理 34.6.)

这样, $\mathscr{G}$ 是充分 $\sigma-$ 代数, 即满足性质 (47). 假设 $\mathbf{P}_\theta \ll \mathbf{P}_{\theta_0}, \theta \in \Theta$, 证明, 对于每一个 $\theta \in \Theta, g_\theta^{(\theta_0)} = \mathrm{d}\mathbf{P}_\theta / \mathrm{d}\mathbf{P}_{\theta_0}$ 是 $\mathscr{G}-$ 可测函数.

设 $A \in \mathscr{F}$. 那么, 对于每一个 $\theta \in \Theta$, 利用条件数学期望的性质, 可得 (其中 $g_\theta^{(\theta_0)} = \mathrm{d}\mathbf{P}_\theta/\mathrm{d}\mathbf{P}_{\theta_0}$):

$$\begin{aligned}\mathbf{P}_\theta(A) &= \mathbf{E}_\theta I_A = \mathbf{E}_\theta \mathbf{E}_\theta(I_A|\mathscr{G}) = \mathbf{E}_\theta \mathbf{E}_{\theta_0}(I_A|\mathscr{G}) = \mathbf{E}_{\theta_0}[g_\theta^{(\theta_0)} \mathbf{E}_{\theta_0}(I_A|\mathscr{G})] \\ &= \mathbf{E}_{\theta_0}\mathbf{E}_{\theta_0}[g_\theta^{(\theta_0)}\mathbf{E}_{\theta_0}(I_A|\mathscr{G})|\mathscr{G}] = \mathbf{E}_{\theta_0}\Big([\mathbf{E}_{\theta_0}(g_\theta^{(\theta_0)}|\mathscr{G})] \times [\mathbf{E}_{\theta_0}(I_A|\mathscr{G})]\Big) \\ &= \mathbf{E}_{\theta_0}\mathbf{E}_{\theta_0}[I_A \mathbf{E}_{\theta_0}(g_\theta^{(\theta_0)}|\mathscr{G})|\mathscr{G}] = \mathbf{E}_{\theta_0} I_A \mathbf{E}_{\theta_0}(g_\theta^{(\theta_0)}|\mathscr{G}) = \int_A \mathbf{E}_{\theta_0}(g_\theta^{(\theta_0)}|\mathscr{G})\mathrm{d}\mathbf{P}_{\theta_0}.\end{aligned}$$

从而, 导数的变式 $g_\theta^{(\theta_0)} = \mathrm{d}\mathbf{P}_\theta/\mathrm{d}\mathbf{P}_{\theta_0}$ 是 $\mathscr{G}-$ 可测函数 $\mathbf{E}_{\theta_0}(g_\theta^{(\theta_0)}|\mathscr{G})$.

这样, 当 $\lambda = \mathbf{P}_{\theta_0}$ 时, 由 $\sigma-$ 代数 $\mathscr{G}$ 的充分性, 可得因子分解的性质 (48), 其中 $\widehat{g}_\theta^{(\theta_0)}(\omega) = g_\theta^{(\theta_0)}, h \equiv 1$.

对于一般情形 (仍然需要补充假设 $\mathbf{P}_\theta \ll \mathbf{P}_{\theta_0}, \theta, \in \Theta$), 有

$$g_\theta^{(\lambda)} = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\lambda} = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mathbf{P}_{\theta_0}} \times \frac{\mathrm{d}\mathbf{P}_{\theta_0}}{\mathrm{d}\lambda} = g_\theta^{(\theta_0)}\frac{\mathrm{d}\mathbf{P}_{\theta_0}}{\mathrm{d}\lambda}.$$

记

$$\widehat{g}_\theta^{(\theta_0)} = g_\theta^{(\theta_0)}, \quad h = \frac{\mathrm{d}\mathbf{P}_{\theta_0}}{\mathrm{d}\lambda},$$

得所要证明的因子分解表达式 (48). □

注 3 有益地强调, 对于任意测度族 $\mathscr{P} = \{\mathbf{P}_\theta, \theta \in \Theta\}$, (在没有任何控制型假设的情况下) 充分 $\sigma-$ 代数显然存在. 作为这样的 $\sigma-$ 代数, 可以取最 "丰富" 的 $\sigma-$ 代数 $\mathscr{F}$.

事实上, 这时对于每一个可积随机变量 $X, \mathbf{E}_\theta(X|\mathscr{F}) = X(\mathbf{P}_\theta - \text{a.c.})$, 从而性质 (47) 成立.

显然, 这样的充分 $\sigma-$ 代数没有什么意义, 因为这时不会有任何 "数据的压缩". 寻找最小 $\sigma-$ 代数 $\mathscr{G}_{\min}$, 即一切充分 $\sigma-$ 代数的交才有真正的意义 (参见 §2 的引理 1 的证明, 可见最小 $\sigma-$ 代数存在). 然而明显地建立这样的 $\sigma-$ 代数并不容易 (不过可以参见 [107] 第二章 §13 ~ §15).

注 4 假设 $\mathscr{P} = \{\mathbf{P}_\theta, \theta \in \Theta\}$ 是被控制族 ($\mathbf{P}_\theta \ll \lambda, \theta \in \Theta, \lambda$ 是 $\sigma-$ 有限测度), 而对于一切 $\theta \in \Theta$, 密度 $g_\theta^{(\lambda)} = (d\mathbf{P}_\theta/\mathrm{d}\lambda)(\omega)$ 可以表示为

$$g_\theta^{(\lambda)}(\omega) = G_\theta^{(\lambda)}(T(\omega))h(\omega), \quad (\lambda - \text{a.c.}), \tag{53}$$

其中 $T = T(\omega)$ 是取值于集合 E 的 (及在 E 上分出的 $\sigma-$ 代数 $\mathscr{E}$) 某 $\mathscr{F}/\mathscr{E}-$ 可测函数 (随机元, 见 §5). 设函数 $G_\theta^{(\lambda)}(t), t \in E$, 和 $h(\omega), \omega \in \Omega$, 是非负的且相应为 $\mathscr{E}-$ 和 $\mathscr{F}-$ 可测的.

比较 (48) 式和 (53) 式, 可见 $\sigma-$ 代数 $\mathscr{G} = \sigma(T(\omega))$ 是充分的, 而函数 $T = T(\omega)$ (在定义 10 的意义下) 是充分统计量.

注意, 在被控制的情形下, 通常正式以满足因子分解式 (53) 的函数 $T = T(\omega)$ 当作充分统计量的定义.

例 5 (指数族) 假设 $\Omega = \mathbb{R}^n, \mathscr{F} = \mathscr{B}(\mathbb{R}^n)$, 测度 $\mathbf{P}_\theta$ 是按如下方式构造的: 对于 $\omega = (x_1, x_2, \cdots, x_n), x_i \in \mathbb{R}$, 则

$$\mathbf{P}_\theta(\mathrm{d}\omega) = \mathbf{P}_\theta(\mathrm{d}x_1)\cdots\mathbf{P}_\theta(\mathrm{d}x_n), \tag{54}$$

其中测度 $\mathbf{P}_\theta(\mathrm{d}x), x \in \mathbb{R}$, 有如下构造:

$$\mathbf{P}_\theta(\mathrm{d}x) = \alpha(\theta)\mathrm{e}^{\beta(\theta)s(x)}\gamma(x)\lambda(\mathrm{d}x). \tag{55}$$

这里, $s = s(x)$ 是某一 $\mathscr{B}$- 可测函数, 而 $\alpha(\theta), \beta(\theta), \gamma(x), \lambda(\mathrm{d}x)$ 含义明显. (测度族 $\mathbf{P}_\theta, \theta \in \Theta$, 是称为指数族的最简单例子.) 由 (54) 式和 (55) 式, 可见

$$\mathbf{P}_\theta(\mathrm{d}\omega) = \alpha^n(\theta)\mathrm{e}^{\beta(\theta)[s(x_1)+\cdots+s(x_n)]}\gamma(x_1)\cdots\gamma(x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n. \tag{56}$$

由 (56) 式与 (53) 式比较, 可见统计量 $T(\omega) = s(x_1)+\cdots+s(x_n), \omega = (x_1, \cdots, x_n)$, (对于所考虑的指数族) 是充分统计量.

如果对于 $\omega = (x_1, \cdots, x_n)$, 记 $X_1(\omega) = x_1, \cdots, X_n(\omega) = x_n$, 则 (按测度的直积原则组成的) 测度 $\mathbf{P}_\theta$ 的构造, 使 $X_1, \cdots, X_n$ 关于这些测度是独立同分布随机变量序列.

于是, 根据序列 $X_1(\omega), \cdots, X_n(\omega)$ 建立的统计量 $T(\omega) = s(X_1(\omega)) + \cdots + s(X_n(\omega))$, 是充分统计量. (在练习题 20 中要求回答, 这一统计量是否最小充分统计量).

例 6 设 $\Omega = \mathbb{R}^n, \mathscr{F} = \mathscr{B}(\mathbb{R}^n)$, 参数 $\theta > 0$ 且对于 $\omega = (x_1, \cdots, x_n)$, (关于勒贝格测度 λ 的) 密度为

$$\frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\lambda}(\omega) = \begin{cases} \theta^{-n}, & \text{若 } 0 \leqslant x_i \leqslant \theta, i = 1, \cdots n, \\ 0, & \text{若不然.} \end{cases}$$

如果

$$T(\omega) = \max_{1 \leqslant i \leqslant n} x_i,$$

$$h(\omega) = \begin{cases} 1, & \text{若 } x_i \geqslant 0, i = 1, \cdots, n, \\ 0, & \text{若不然,} \end{cases}$$

$$G_\theta^{(\lambda)}(t) = \begin{cases} \theta^{-n}, & \text{若 } 0 \leqslant t \leqslant \theta, \\ 0, & \text{若不然,} \end{cases}$$

则

$$\frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\lambda}(\omega) = G_\theta^{(\lambda)}(T(\omega))h(\omega). \tag{57}$$

于是, $T(\omega) = \max\limits_{1 \leqslant i \leqslant n} x_i$ 是充分统计量得证.

11. 无偏估计量, 拉奥 – 布莱克韦尔 (C. R. Rao-D. H. Blackwell) 定理 假设 Θ 是直线上的一子集, $\mathscr{E}=(\Omega,\mathscr{F},\mathscr{P}=\{\mathbf{P}_\theta:\theta\in\Theta\})$ 是概率统计模型. 现在考虑的问题是: 建立参数 θ 的 "好" 估计量.

任何随机变量 $\widehat{\theta}=\widehat{\theta}(\omega)$ 都将视为参数 θ 的估计量 (对照第一章 §7).

下面的定理表明, 充分 $\sigma-$ 代数的概念可以改进估计量的 "质量". 估计量 $\widehat{\theta}$ 的质量, 是用其对参数 θ 的真值的均方偏差来衡量的. 更确切地说, 称参数 θ 的估计量 $\widehat{\theta}$ 是无偏的, 如果 $\mathbf{E}_\theta|\widehat{\theta}|<\infty$ 且 $\mathbf{E}_\theta\widehat{\theta}=\theta$ (对照第一章 §7 性质 2).

定理 8 (拉奥和布莱克韦尔) *设 $\mathscr{G}$ 对于概率测度族 $\mathscr{P}$ 是充分 $\sigma-$ 代数, 而 $\widehat{\theta}=\widehat{\theta}(\omega)$ 是参数 θ 的某一估计量.*

a) *如果 $\widehat{\theta}$ 是无偏估计量, 则估计量*

$$T=\mathbf{E}_\theta(\widehat{\theta}|\mathscr{G}) \tag{58}$$

也是无偏的.

b) *估计量 T 在如下意义上优于估计量 $\widehat{\theta}$:*

$$\mathbf{E}_\theta(T-\theta)^2\leqslant\mathbf{E}_\theta(\widehat{\theta}-\theta)^2,\quad \theta\in\Theta. \tag{59}$$

证明 性质 a) 由下列等式可见:

$$\mathbf{E}_\theta T=\mathbf{E}_\theta\mathbf{E}_\theta(\widehat{\theta}|\mathscr{G})=\mathbf{E}_\theta\widehat{\theta}=\theta.$$

为证明性质 b) 只需注意到: 由延森不等式 (只需在练习题 5 中设 $g(x)=(x-\theta)^2$), 有

$$[\mathbf{E}_\theta(\widehat{\theta}|\mathscr{G})-\theta)]^2\leqslant\mathbf{E}_\theta[(\widehat{\theta}-\theta)^2|\mathscr{G}].$$

在不等式两侧同时求数学期望 $\mathbf{E}_\theta(\cdot)$, 得要证明的 (59) 式. □

12. 练习题

1. 设 ξ 和 η 是独立同分布随机变量, $\mathbf{E}\xi$ 有定义. 证明

$$\mathbf{E}(\xi|\xi+\eta)=\mathbf{E}(\eta|\xi+\eta)=\frac{\xi+\eta}{2}(\text{a.c.}).$$

2. 设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量, 且 $\mathbf{E}|\xi_i|<\infty$. 证明

$$\mathbf{E}(\xi_1|S_n,S_{n+1},\cdots)=\frac{S_n}{n}(\text{a.c.}),$$

其中 $S_n=\xi_1+\cdots+\xi_n$.

3. 假设对于随机元 (X,Y), 存在正则分布 $P_x(B)=\mathbf{P}(Y\in B|X=x)$. 证明, 如果 $\mathbf{E}|g(X,Y)|<\infty$, 则 $P_x-\text{a.c.}$, 有

$$\mathbf{E}[g(X,Y)|X=x]=\int g(x,y)P_x(\mathrm{d}y).$$

4. 设随机变量 ξ 的分布函数为 $F_\xi(x)$, 证明

$$\mathbf{E}(\xi|a<\xi\leqslant b)=\frac{\displaystyle\int_a^b x\mathrm{d}F_\xi(x)}{F_\xi(b)-F_\xi(a)}$$

(假设 $F_\xi(b)-F_\xi(a)>0$).

5. 设 $g=g(x)$ 是凹 (向下凸) 博雷尔函数, $\mathbf{E}|g(\xi)|<\infty$. 证明, 对于条件数学期望, (a.c.) 有延森不等式:

$$g[\mathbf{E}(\xi|\mathscr{G})]\leqslant\mathbf{E}[g(\xi)|\mathscr{G}].$$

6. 证明随机变量 ξ 和 σ - 代数 $\mathscr{G}$ 独立 (即对于任意 $B\in\mathscr{G}$, 随机变量 ξ 与 I_B 独立), 当且仅当对于每一博雷尔函数 $g(x), \mathbf{E}|g(\xi)|<\infty$, 有 $\mathbf{E}[g(\xi)|\mathscr{G}]=\mathbf{E}g(\xi)$.

7. 设 ξ 是非负随机变量, $\mathscr{G}$ 是 σ - 代数 $\mathscr{G}\subseteq\mathscr{F}$, $\mathbf{Q}$ 是由等式 $\mathbf{Q}(A)=\int_A\xi\mathrm{d}\mathbf{P}$ 定义在集合 $A\in\mathscr{G}$ 上的测度. 证明, $\mathbf{E}(\xi|\mathscr{G})<\infty$ (a.c.) 的充分和必要条件为 $\mathbf{Q}$ 是有限测度.

8. 证明条件概率 $\mathbf{P}(A|B)$ 在如下意义上 "连续": 如果 $\lim_n A_n=A, \lim_n B_n=B$ 且 $\mathbf{P}(B_n)>0, \mathbf{P}(B)>0$, 则 $\lim_n\mathbf{P}(A_n|B_n)=\mathbf{P}(A|B)$.

9. 设 $\Omega=(0,1), \mathscr{F}=\mathscr{B}((0,1)), \mathbf{P}$ 是勒贝格测度; $X(\omega)$ 和 $Y(\omega)$ 是在 (0,1) 上均匀分布的两个独立随机变量. 考虑第三个随机变量 $Z(\omega)=|X(\omega)-Y(\omega)|$ 是 "点" $X(\omega)$ 和 $Y(\omega)$ 之间的距离. 证明 $Z(\omega)$ 的分布函数 $F_Z(z)$ 有密度:

$$f_Z(z)=\begin{cases}2(1-z), & \text{若 } 0\leqslant z\leqslant 1,\\ 0, & \text{若不然}.\end{cases}$$

(由此可见 $\mathbf{E}Z=1/3$.)

10. 向半径为 R 的圆 $\{(x,y):x^2+y^2\leqslant R^2\}$ 上 "随机地" 掷两点 A_1 和 A_2, 即独立地选两点且选到点 $A_i(\rho_i,\theta_i), i=1,2$, 的概率 (在极坐标系中) 为:

$$\mathbf{P}\{\rho_i\in\mathrm{d}r,\ \theta_i\in\mathrm{d}\theta\}=\frac{r\mathrm{d}r\mathrm{d}\theta}{\pi R^2},\quad i=1,2.$$

证明点 A_1 和 A_2 之间的距离 ρ 的分布密度为:

$$f_\rho(r)=\frac{2r}{\pi R^2}\left[2\arccos\left(\frac{r}{2R}\right)-\frac{r}{R}\sqrt{1-\left(\frac{r}{2R}\right)^2}\right],$$

其中 $0<r<2R$.

11. 在单位正方形中 (其顶点为 (0,0),(0,1),(1,1),(1,0)) "随机地" (解释何意!) 选一点 $P=(x,y)$. 证明该点 P 离点 (1,1) 比离点 $(-1/2,1/2)$ 更近.

12. A 和 B 两人约好在 7 点到 8 点之间会面. 但是两人忘记了确切的会面时间, 只好在 7 点到 8 点之间 "随机地" 到达会面地点, 并且最多等 10 分钟. 证明他们能见面的概率等于 11/36.

13. 设 $X_1, X_2, \cdots$ 是独立随机变量序列, $S_n = \sum_{i=1}^n X_i$, 而 $\sigma(S_2)$ 是由 S_2 生成的 σ - 代数. 证明 S_1 和 S_3 关于 $\sigma(S_2)$ 条件独立.

14. 称 σ - 代数 $\mathscr{G}_1$ 和 $\mathscr{G}_2$ 关于 σ - 代数 $\mathscr{G}_3$ 条件独立, 如果对于任意 $A_i \in \mathscr{G}_i, i = 1, 2$,

$$\mathbf{P}(A_1 A_2 | \mathscr{G}_3) = \mathbf{P}(A_1 | \mathscr{G}_3) \mathbf{P}(A_2 | \mathscr{G}_3).$$

证明 $\mathscr{G}_1$ 和 $\mathscr{G}_2$ 关于 $\mathscr{G}_3$ 条件独立, 与下列任何条件之一等价 ($\mathbf{P}$ - a.c.):

(a) 对于一切 $A_1 \in \mathscr{G}_1$, 有 $\mathbf{P}(A_1 | \sigma(\mathscr{G}_2 \bigcup \mathscr{G}_3)) = \mathbf{P}(A_1 | \mathscr{G}_3)$;

(b) 对于一切集合 $B \in \mathscr{P}_1$, 其中集系 $\mathscr{P}_1$ 形成 π- 系 $\mathscr{G}_1 = \sigma(\mathscr{P}_1)$, 有 $\mathbf{P}(B | \sigma(\mathscr{G}_2 \bigcup \mathscr{G}_3)) = \mathbf{P}(B | \mathscr{G}_3)$;

(c) 对于一切分别属于 π - 系 $\mathscr{P}_1$ 和 $\mathscr{P}_2$ 的集合 B_1 和 B_2, 而且其中 $\mathscr{G}_1 = \sigma(\mathscr{P}_1)$ 和 $\mathscr{G}_2 = \sigma(\mathscr{P}_2)$, 有 $\mathbf{P}(B_1 B_2 | \sigma(\mathscr{G}_2 \bigcup \mathscr{G}_3)) = \mathbf{P}(B_1 | \mathscr{G}_3) \mathbf{P}(B_2 | \mathscr{G}_3)$;

(d) 对于一切 $\sigma(\mathscr{G}_2 \bigcup \mathscr{G}_3)$ - 可测且有数学期望 $\mathbf{E}X$ 的随机变量 X (见 §6 定义 2), 有 $\mathbf{E}(X | \sigma(\mathscr{G}_2 \bigcup \mathscr{G}_3)) = \mathbf{E}(X | \mathscr{G}_3)$.

15. 证明关于条件数学期望的法图引理的如下扩展变式 (对照定理 2 之 (d)).

设概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$, 而 $(\xi_n)_{n \geqslant 1}$ 是随机变量序列, 且数学期望 $\mathbf{E}\xi_n, n \geqslant 1$ 而 $\mathbf{E}\underline{\lim}\xi_n$ 存在 (亦可取 $\pm\infty$ 为值; 见 §6 定义 2).

假设 $\mathscr{G}$ 是 $\mathscr{F}$ 中事件的子 σ - 代数, 而且当 $a \to \infty$ 时,

$$\sup_{n \geqslant 1} \mathbf{E}(\xi_n^- I_{\{\xi_n \geqslant a\}} | \mathscr{G}) \to 0, \quad (\mathbf{P} - \text{a.c.}),$$

证明

$$\mathbf{E}(\underline{\lim}\xi_n | \mathscr{G}) \leqslant \underline{\lim}\mathbf{E}(\xi_n | \mathscr{G}), \quad (\mathbf{P} - \text{a.c.}).$$

16. 假设像上题一样, 对于随机变量序列 $(\xi_n)_{n \geqslant 1}$, 数学期望 $\mathbf{E}\xi_n, n \geqslant 1$, 存在, 而 $\mathscr{G}$ 是 $\mathscr{F}$ 中事件的子 σ - 代数, 且

$$\sup_n \lim_{k \to \infty} \mathbf{E}(|\xi_n| I_{\{|\xi_n| \geqslant k\}} | \mathscr{G}) = 0, \quad (\mathbf{P} - \text{a.c.}). \tag{60}$$

证明, 如果 $\xi_n \to \xi(\mathbf{P} - \text{a.c.})$, 且数学期望 $\mathbf{E}\xi_n, n \geqslant 1$, 存在, 则

$$\mathbf{E}(\xi_n | \mathscr{G}) \to \mathbf{E}(\xi | \mathscr{G}), \quad (\mathbf{P} - \text{a.c.}).$$

17. 假设在上题的条件下, 式 (60) 换成满足条件: 对于某一 $\alpha > 1$,

$$\sup_n \mathbf{E}(|\xi_n|^\alpha | \mathscr{G}) < \infty, \quad (\mathbf{P} - \text{a.c.}).$$

那么,

$$\mathbf{E}(\xi_n | \mathscr{G}) \to \mathbf{E}(\xi | \mathscr{G}), \quad (\mathbf{P} - \text{a.c.}).$$

18. 设对于某个 $p \geqslant 1, \xi_n \xrightarrow{L^p} \xi$, 证明 $\mathbf{E}(\xi_n | \mathscr{G}) \xrightarrow{L^p} \mathbf{E}(\xi | \mathscr{G})$.

19. (a) 设 $\mathbf{D}(X|Y) \equiv \mathbf{E}\{[X - \mathbf{E}(X|Y)]^2|Y\}$, 则 $\mathbf{D}X = \mathbf{ED}(X|Y) + \mathbf{DE}(X|Y)$.
(b) 证明 $\operatorname{cov}(X, Y) = \operatorname{cov}(X, \mathbf{E}(Y|X))$.
20. 说明例 5 中的充分统计量 $T(\omega) = s(X_1(\omega)) + \cdots + s(X_n(\omega))$ 是否最小的.
21. 证明因子分解公式 (57).
22. 对于第 10 小节的例 5, 证明

$$\mathbf{E}_\theta(X_i|T) = \frac{n+1}{2n}T,$$

其中对于 $\omega = (x_1, \cdots, x_n), X_i(\omega) = x_i, i = 1, \cdots, n$.

§8. 随机变量 II

1. 方差, 协方差和相关系数 在第一章引进了简单随机变量的一些简单数字特征: 如方差、协方差和相关系数. 在一般情形下也相应地引进这些概念. 具体地说, 假设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, 而随机变量 $\xi = \xi(\omega)$ 有数学期望 $\mathbf{E}\xi$.

称

$$\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2$$

为随机变量 ξ 的*方差*.

称 $\sigma = +\sqrt{\mathbf{D}\xi}$ 为随机变量 ξ 的*标准差*.

如果随机变量 ξ 有高斯 (正态) 密度:

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \sigma > 0, \quad -\infty < x < +\infty, \tag{1}$$

则 (1) 式中参数的含义非常简单:

$$m = \mathbf{E}\xi, \quad \sigma^2 = \mathbf{D}\xi.$$

这样, 称为*高斯分布*或*正态分布*的随机变量 ξ 的密度, 完全决定于其均值 m 和方差 σ^2. (因此, 对于参数为 m 和 σ^2 的正态分布, 常使用记号: $\xi \sim N(m, \sigma^2)$.)

现在假设 (ξ, η) 是随机向量, 称

$$\operatorname{cov}(\xi, \eta) = \mathbf{E}(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta) \tag{2}$$

为 ξ 和 η 的*协方差* (假设其中的各个数学期望存在).

如果 $\operatorname{cov}(\xi, \eta) = 0$, 则称随机变量 ξ 和 η *不相关*.

如果 $0 < \mathbf{D}\xi < \infty, 0 < \mathbf{D}\eta < \infty$, 则称

$$\rho(\xi, \eta) = \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{\mathbf{D}\xi \times \mathbf{D}\eta}} \tag{3}$$

为随机变量 ξ 和 η 的*相关系数*.

在第一章 §4 曾讲述了简单随机变量的方差、协方差和相关系数的性质. 在一般情形下, 这些性质的提法完全类似.

设随机向量 $\xi=(\xi_1,\cdots,\xi_n)$ 的分量有有限二阶矩. 称 $n\times n$ 阶矩阵 $\boldsymbol{R}=(R_{ij})_{n\times n}$ 为随机向量 (ξ,η) 的*协方差矩阵*, 其中 $R_{ij}=\mathrm{cov}(\xi_i,\xi_j)$. 显然, $\boldsymbol{R}$ 是对称矩阵. 此外, 由于对于任意 $\lambda_i\in\mathbb{R}, i=1,\cdots,n$,

$$\sum_{i,j=1}^{n} R_{ij}\lambda_i\lambda_j=\mathbf{E}\left[\sum_{i=1}^{n}(\xi_i-\mathbf{E}\xi_i)\lambda_i\right]^2\geqslant 0,$$

可见 $\boldsymbol{R}$ 是*非负定矩阵*, 即

$$\sum_{i,j=1}^{n} R_{ij}\lambda_i\lambda_j\geqslant 0.$$

下面的引理说明逆命题亦成立.

引理 *$n\times n$ 阶矩阵 $\boldsymbol{R}=(R_{ij})_{n\times n}$ 为随机向量 (ξ,η) 的协方差矩阵的充分和必要条件是, $\mathbb{R}$ 为对称和非负定矩阵, 或者等价地: 存在 $n\times k(1\leqslant k\leqslant n)$ 阶矩阵 $\boldsymbol{A}$, 使*

$$\boldsymbol{R}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}},$$

其中上标 "T" 是矩阵的 "转置" 符号: $\boldsymbol{A}^{\mathrm{T}}$ 是矩阵 $\boldsymbol{A}$ 的转置.

证明 上面已经证明, 任何协方差矩阵都是对称和非负定矩阵.

反之, 设 $\boldsymbol{R}$ 是对称的非负定矩阵. 由矩阵理论知, 对于任意对称非负定矩阵, 存在正交矩阵 $\boldsymbol{Z}$ (即 $\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}=\boldsymbol{E}$ 是单位矩阵), 使

$$\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{Z}=\boldsymbol{D},$$

其中 $\boldsymbol{D}$ 是对角线上元素为非负 $(d_i\geqslant 0, i=1,\cdots,n)$ 的对角矩阵:

$$\boldsymbol{D}=\begin{pmatrix} d_1 & 0 & \cdots & 0\\ 0 & d_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & d_n\end{pmatrix}.$$

由此可见,

$$\boldsymbol{R}=\boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^{\mathrm{T}}=(\boldsymbol{Z}\boldsymbol{B})(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Z}^{\mathrm{T}}),$$

其中 $\boldsymbol{B}$ 是对角线上元素为 $b_i=+\sqrt{d_i}(i=1,\cdots,n)$ 的对角矩阵. 因此, 如果设 $\boldsymbol{A}=\boldsymbol{Z}\boldsymbol{B}$, 则对于 $\boldsymbol{R}$, 得所要求的表现 $\boldsymbol{R}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.

显然, $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 就是所要求的对称非负定矩阵. 因此, 只需证明 $\boldsymbol{R}$ 是某随机向量的协方差矩阵.

设 $\eta_1, \eta_2, \cdots, \eta_n$ 是独立同正态分布 $N(0,1)$ 的随机变量序列. (由 §9 定理 1 之系 1, 可见这样的随机变量序列存在; 实际上, 这容易由 §3 定理 2 得到) 那么, 随机向量 $\xi = \boldsymbol{A}\eta$ 具有所要求的性质 (注意, $\xi = \boldsymbol{A}\eta$ 应视为列向量). 事实上,

$$\mathbf{E}\xi\xi^{\mathrm{T}} = \mathbf{E}(\boldsymbol{A}\eta)(\boldsymbol{A}\eta)^{\mathrm{T}} = \boldsymbol{A}(\mathbf{E}\eta\eta^{\mathrm{T}})\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{E}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}.$$

(如果 $\zeta = (\zeta_{ij})$ 是以随机变量为元素的矩阵, 则 $\mathbf{E}\zeta$ 表示 $(\mathbf{E}\zeta_{ij})$.) □

现在考虑二维高斯 (正态) 密度:

$$f_{\xi,\eta}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2} -2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}, \tag{4}$$

其中含 5 个参数: $m_1, m_2, \sigma_1, \sigma_2, \rho$ (对照 §3 之 (14) 式), 其中 $|m_1| < \infty, |m_2| < \infty, \sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$. (见 §3 图 28.) 经过简单的计算, 可以说明这些参数的概率意义:

$$\begin{aligned} m_1 = \mathbf{E}\xi, \quad \sigma_1^2 = \mathbf{D}\xi, \\ m_2 = \mathbf{E}\eta, \quad \sigma_2^2 = \mathbf{D}\eta, \\ \rho = \rho(\xi,\eta). \end{aligned}$$

在第一章 §4 曾经说明, 如果随机变量 ξ 和 η 不相关, 则还不能说明它们独立. 但是, 假如 (ξ,η) 是正态的 (高斯的), 则由 ξ 和 η 不相关, 可见它们独立.

事实上, 如果在 (4) 式中 $\rho = 0$, 则

$$f_{\xi,\eta}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{(x-m_1)^2}{2\sigma_1^2}\right\}\exp\left\{-\frac{(y-m_2)^2}{2\sigma_2^2}\right\}.$$

但是, 由于 §6 的 (55) 式和本小节 (4) 式, 有

$$\begin{aligned} f_\xi(x) &= \int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\mathrm{d}y = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-m_1)^2}{2\sigma_1^2}\right\}, \\ f_\eta(y) &= \int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(y-m_2)^2}{2\sigma_2^2}\right\}. \end{aligned}$$

于是,

$$f_{\xi,\eta}(x,y) = f_\xi(x)f_\eta(y),$$

由此可见, 随机变量 ξ 和 η 独立 (见 §6 的第 9 小节末尾).

2. 最优估计量 在前面 §7 引进的条件数学期望概念, 其有说服力的重要性在于它在解决关于参数估计理论的如下问题中的应用 (对照第一章 §4 的第 8 小节).

设 (ξ,η) 是随机向量, 其中 ξ 可以观测, 而 η 不可观测. 问如何根据对 ξ 的观测值, "估计" 不可观测的分量 η?

为使问题更加明确, 我们引进**估计量**的概念. 设 $\varphi=\varphi(x)$ 是博雷尔函数. 称随机变量 $\varphi(\xi)$ 为由 ξ 对 η 的**估计量**, 而称 $\mathbf{E}[\eta-\varphi(\xi)]^2$ 为该估计量的 (**均根方**) **误差**. 估计量 $\varphi^*(\xi)$ 称做 (在根均方意义上) **最优的**, 如果

$$\Delta \equiv \mathbf{E}[\eta-\varphi^*(\xi)]^2=\inf_{\varphi}\mathbf{E}[\eta-\varphi(\xi)]^2, \tag{5}$$

其中 inf 在一切博雷尔函数 $\varphi=\varphi(x)$ 类中来求.

定理 1 *设 $\mathbf{E}\eta^2<\infty$. 那么, 最优估计量 $\varphi^*(x)$ 存在, 并且作为 $\varphi^*(\xi)$ 可以选任意函数:*

$$\varphi^*(x)=\mathbf{E}(\eta|\xi=x). \tag{6}$$

证明 不失普遍性, 可以只考虑满足 $\mathbf{E}\varphi^2(\xi)<\infty$ 的 $\varphi(\xi)$. 那么, 如果 $\varphi(\xi)$ 是所欲求的估计量, 而 $\varphi^*(\xi)=\mathbf{E}(\eta|\xi)$, 则

$$\begin{aligned}\mathbf{E}[\eta-\varphi(\xi)]^2&=\mathbf{E}\{[\eta-\varphi^*(\xi)]+[\varphi^*(\xi)-\varphi(\xi)]\}^2\\&=\mathbf{E}[\eta-\varphi^*(\xi)]^2+\mathbf{E}[\varphi^*(\xi)-\varphi(\xi)]^2+2\mathbf{E}[\eta-\varphi^*(\xi)][\varphi^*(\xi)-\varphi(\xi)]\\&\geqslant \mathbf{E}[\eta-\varphi^*(\xi)]^2,\end{aligned}$$

由于 $\mathbf{E}[\varphi^*(\xi)-\varphi(\xi)]^2\geqslant 0$, 并且由条件数学期望的性质, 可见

$$\begin{aligned}\mathbf{E}[\eta-\varphi^*(\xi)][\varphi^*(\xi)-\varphi(\xi)]&=\mathbf{E}\{\mathbf{E}[\eta-\varphi^*(\xi)][\varphi^*(\xi)-\varphi(\xi)]|\xi\}\\&=\mathbf{E}\{[\varphi^*(\xi)-\varphi(\xi)]\mathbf{E}[\eta-\varphi^*(\xi)|\xi]\}=0.\end{aligned}$$ □

注 1 由定理的证明, 可见如果 ξ 不是随机变量, 而是在某可测空间 $(E,\mathscr{E})$ 取值的随机元, 则其结论仍然成立. 在后一种情形下, 估计量 $\varphi=\varphi(x)$ 应理解为 $\mathscr{E}/\mathscr{B}(\mathbb{R})$ - 可测函数.

现在, 假设 (ξ,η) 是服从高斯分布 (正态分布) 的随机向量, 并讨论 $\varphi^*(x)$ 的结构, 其中 (4) 式为二维正态分布的密度.

由 §7 中 (1),(4) 和 (18) 各式知, 条件概率分布密度 $f_{\eta|\xi}(y|x)$ 为

$$f_{\eta|\xi}(y|x)=\frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_2^2}}\exp\left\{-\frac{(y-m(x))^2}{2\sigma_2^2(1-\rho^2)}\right\}, \tag{7}$$

其中

$$m(x)=m_2+\frac{\sigma_2}{\sigma_1}\rho(x-m_1). \tag{8}$$

那么, 由 §7 定理 3 的系, 可见

$$\mathbf{E}(\eta|\xi=x)=\int_{-\infty}^{\infty}yf_{\eta|\xi}(y|x)\mathrm{d}y=m(x), \tag{9}$$

且

$$\mathbf{D}(\eta|\xi=x) \equiv \mathbf{E}\{[\eta-\mathbf{E}(\eta|\xi=x)]^2|\xi=x\}$$
$$=\int_{-\infty}^{\infty}[y-m(x)]^2 f_{\eta|\xi}(y|x)\mathrm{d}y=\sigma_2^2(1-\rho^2). \tag{10}$$

注意. 条件方差 $\mathbf{D}(\eta|\xi=x)$ 不依赖于 x, 因此,

$$\Delta=\mathbf{E}[\eta-\mathbf{E}(\eta|\xi=x)]^2=\sigma_2^2(1-\rho^2). \tag{11}$$

公式 (9) 和 (11) 是在假设 $\mathbf{D}\xi>0, \mathbf{D}\eta>0$ 条件下得到的. 假如 $\mathbf{D}\xi>0$, 而 $\mathbf{D}\eta=0$, 则 (9) 和 (11) 两式显然成立.

于是, 得到下面的结果 (对照第一章 §4 的 (16) 式和 (17) 式).

定理 2 (正态相关定理) *设 (ξ,η) 是 $\mathbf{D}\xi>0$ 的高斯 (正态) 随机向量. 由 ξ 对 η 的最优估计量是:*

$$\mathbf{E}(\eta|\xi)=\mathbf{E}\eta+\frac{\operatorname{cov}(\xi,\eta)}{\mathbf{D}\xi}(\xi-\mathbf{E}\xi), \tag{12}$$

而其误差为

$$\Delta \equiv \mathbf{E}[\eta-\mathbf{E}(\eta|\xi)]^2=\mathbf{D}\eta-\frac{\operatorname{cov}^2(\xi,\eta)}{\mathbf{D}\xi}. \tag{13}$$

注 2 曲线 $y(x)=\mathbf{E}(\eta|\xi=x)$ 称做 *η 在 ξ 上的或 η 关于 ξ 的回归曲线*. 在正态情形下 $\mathbf{E}(\eta|\xi=x)=ax+b$, 因而 η 关于 ξ 的回归是*线性的*. 因此第一章 §4 中, 对于最优线性估计及其误差, (12) 和 (13) 式的右侧与 (16) 和 (17) 式的右侧对应相等, 就毫不意外.

系 *设 ε_1 和 ε_2 是均值为 0 和方差为 1 的独立正态随机变量, 而*

$$\xi=a_1\varepsilon_1+a_2\varepsilon_2, \quad \eta=b_1\varepsilon_1+b_2\varepsilon_2.$$

那么, $\mathbf{E}\xi=\mathbf{E}\eta=0, \mathbf{D}\xi=a_1^2+a_2^2, \mathbf{D}\eta=b_1^2+b_2^2, \operatorname{cov}(\xi,\eta)=a_1b_1+a_2b_2$, 而且若 $a_1^2+a_2^2>0$, 则

$$\mathbf{E}(\eta|\xi)=\frac{a_1b_1+a_2b_2}{a_1^2+a_2^2}\xi, \tag{14}$$

$$\Delta=\frac{(a_1b_2-a_2b_1)^2}{a_1^2+a_2^2}. \tag{15}$$

3. 随机变量的函数的分布 假设一个随机变量是另一个随机变量的函数. 现在, 讨论求随机变量函数的概率分布的问题.

设随机变量 ξ 的分布函数为 $F_\xi(x)$ (如果它有密度, 则记作 $f_\xi(x)$), $\varphi=\varphi(x)$ 是某一博雷尔函数, 而 $\eta=\varphi(\xi)$. 记 $I_y=(-\infty,y]$, 则

$$F_\eta(y)=\mathbf{P}\{\eta \leqslant y\}=\mathbf{P}\{\varphi(\xi)\in I_y\}=\mathbf{P}\{\xi\in\varphi^{-1}(I_y)\}=\int_{\varphi^{-1}(I_y)}F_\xi(\mathrm{d}x). \tag{16}$$

于是, 随机变量 $\eta=\varphi(\xi)$ 的分布函数 $F_\eta(y)$, 可以通过随机变量 ξ 的分布函数 $F_\xi(x)$ 和函数 φ 来表示.

例如, 若 $\eta=a\xi+b, a>0$, 则

$$F_\eta(y)=\mathbf{P}\left\{\xi\leqslant\frac{y-b}{a}\right\}=F_\xi\left(\frac{y-b}{a}\right). \tag{17}$$

设 $\eta=\xi^2$, 则对于 $y<0$, 显然 $F_\eta(y)=0$, 而对于 $y\geqslant 0$, 有

$$\begin{aligned}F_\eta(y)&=\mathbf{P}\{\xi^2\leqslant y\}=\mathbf{P}\{-\sqrt{y}\leqslant\xi\leqslant\sqrt{y}\}\\&=F_\xi(\sqrt{y})-F_\xi(-\sqrt{y})+\mathbf{P}\{\xi=-\sqrt{y}\}.\end{aligned} \tag{18}$$

现在考虑求密度 $f_\eta(y)$ 的问题.

假设随机变量 ξ 的值域是 (有限或无限) 开区间 $I=(a,b)$, 而函数 $\varphi=\varphi(x)$ 对于 $x\in I$ 有定义、连续可微, 并且是单调的: 或严格单调增加, 或严格单调减小. 此外, 假设 $\varphi'(x)\neq 0, x\in I$.

记 $h(y)=\varphi^{-1}(y)$, 并且不失普遍性假设 $\varphi(x)$ 严格增加. 那么, 对于 $y\in\{\varphi(x): x\in I\}$, 有

$$\begin{aligned}F_\eta(y)&=\mathbf{P}\{\eta\leqslant y\}=\mathbf{P}\{\varphi(\xi)\leqslant y\}=\mathbf{P}\{\xi\leqslant\varphi^{-1}(y)\}\\&=\mathbf{P}\{\xi\leqslant h(y)\}=\int_{-\infty}^{h(y)}f_\xi(x)\mathrm{d}x.\end{aligned} \tag{19}$$

根据 §6 练习题 15,

$$\int_{-\infty}^{h(y)}f_\xi(x)\mathrm{d}x=\int_{-\infty}^{y}f_\xi(h(z))h'(z)\mathrm{d}z, \tag{20}$$

因此

$$f_\eta(y)=f_\xi(h(y))h'(y). \tag{21}$$

类似地, 如果 $\varphi(x)$ 严格减小, 则

$$f_\eta(y)=f_\xi(h(y))(-h'(y)).$$

于是, 在两种情形下都有

$$f_\eta(y)=f_\xi(h(y))|h'(y)|. \tag{22}$$

例如, 若 $\eta=a\xi+b, a\neq 0$, 则

$$h(y)=\frac{y-b}{a},\quad f_\eta(y)=\frac{1}{|a|}f_\xi\left(\frac{y-b}{a}\right).$$

如果 $\xi\sim N(m,\sigma^2)$, 而 $\eta=\mathrm{e}^\xi$, 则由 (22) 式, 有

$$f_\eta(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma y}\exp\left\{-\dfrac{\ln(y/M)^2}{2\sigma^2}\right\}, & \text{若 } y>0,\\ 0, & \text{若 } y\leqslant 0,\end{cases} \tag{23}$$

其中 $M = \mathrm{e}^m$.

以 (23) 式为密度的概率分布称做对数正态分布.

如果函数 $\varphi = \varphi(x)$ 不是严格增加或严格减小的, 则 (22) 式对于 $\eta = \varphi(\xi)$ 不能用. 不过, 对于许多应用, 其如下推广就完全够用.

设函数 $\varphi = \varphi(x)$ 定义在集合 $\sum_{k=1}^{n}[a_k, b_k]$ 上, 并且在每一个开区间 $I_k = (a_k, b_k)$ 上连续可微, 以及或者严格增加, 或者严格减小, 并且 $\varphi'(x) \neq 0, x \in I_k$. 此外, 假设 $h_k = h_k(y)$ 是 $\varphi(x), x \in I_k$ 的反函数. 那么, 有 (22) 式如下推广:

$$f_\eta(y) = \sum_{k=1}^{n} f_\xi(h_k(y))|h_k'(y)| I_{D_k}(y), \tag{24}$$

其中 D_k 是函数 $h_k(y)$ 的定义域.

例如, 若 $\eta = \xi^2$, 则设 $I_1 = (-\infty, 0), I_2 = (0, \infty)$, 得 $h_1(y) = -\sqrt{y}, h_2(y) = \sqrt{y}$, 因而

$$f_\eta(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}[f_\xi(\sqrt{y}) + f_\xi(-\sqrt{y})], & 若\ y > 0, \\ 0, & 若\ y \leqslant 0. \end{cases} \tag{25}$$

注意, 由于 $\mathbf{P}\{\xi = -\sqrt{y}\} = 0$, 这一结果亦可由 (18) 式得到. 特别, 如果 $\xi \sim N(0,1)$, 则

$$f_{\xi^2}(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi y}} \mathrm{e}^{-\frac{y}{2}}, & 若\ y > 0, \\ 0, & 若\ y \leqslant 0. \end{cases} \tag{26}$$

经不复杂的计算亦可得到:

$$f_{|\xi|}(y) = \begin{cases} f_\xi(y) + f_\xi(-y), & 若\ y > 0, \\ 0, & 若\ y \leqslant 0. \end{cases} \tag{27}$$

$$f_{+\sqrt{|\xi|}}(y) = \begin{cases} 2y[f_\xi(y^2) + f_\xi(-y^2)], & 若\ y > 0, \\ 0, & 若\ y \leqslant 0. \end{cases} \tag{28}$$

4. 随机变量多元函数的分布　现在考虑多个随机变量的函数.

如果 ξ 和 η 是联合分布函数为 $F_{\xi,\eta}(x, y)$ 的随机变量, 而 $\varphi = \varphi(x, y)$ 是某一博雷尔函数, 则立即可以得到 $\zeta = \varphi(\xi, \eta)$ 的分布函数:

$$F_\zeta(z) = \int_{\{(x,y):\varphi(x,y) \leqslant z\}} \mathrm{d}F_{\xi,\eta}(x, y). \tag{29}$$

例如, 如果 $\varphi(x, y) = x + y$, 而 ξ 和 η 独立 (即 $F_{\xi,\eta}(x, y) = F_\xi(x)F_\eta(y)$), 则利用傅比尼定理, 有

$$\begin{aligned} F_\zeta(z) &= \int_{\{(x,y):x+y \leqslant z\}} \mathrm{d}F_\xi(x)\mathrm{d}F_\eta(y) = \int_{\mathbb{R}^2} I_{\{x+y\leqslant z\}}(x, y)\mathrm{d}F_\xi(x)\mathrm{d}F_\eta(y) \\ &= \int_{-\infty}^{\infty} \mathrm{d}F_\xi(x)\left[\int_{-\infty}^{\infty} I_{\{x+y\leqslant z\}}(x, y)\mathrm{d}F_\eta(y)\right] = \int_{-\infty}^{\infty} F_\eta(z - x)\mathrm{d}F_\xi(x). \end{aligned} \tag{30}$$

类似地可得

$$F_\zeta(z)=\int_{-\infty}^{\infty}F_\xi(z-y)\mathrm{d}F_\eta(y). \tag{31}$$

对于两个分布函数 F 和 G, 函数

$$H(z)=\int_{-\infty}^{\infty}F(z-x)\mathrm{d}G(x)$$

通常记作 $F*G$, 并称做 F 和 G 卷积.

这样, 两个随机变量 ξ 与 η 之和 $\zeta=\xi+\eta$ 的分布函数 F_ζ, 是它们分布函数 F_ξ 与 F_η 的卷积:

$$F_\zeta=F_\xi*F_\eta.$$

显然, 这时 $F_\xi*F_\eta=F_\eta*F_\xi$.

现在假设独立随机变量 ξ 和 η 各自有密度 f_ξ 和 f_η. 那么, 仍然由傅比尼定理及 (31) 式, 有

$$\begin{aligned}F_\zeta(z)&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-y}f_\xi(u)\mathrm{d}u\right]f_\eta(y)\mathrm{d}y=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z}f_\xi(u-y)\mathrm{d}u\right]f_\eta(y)\mathrm{d}y\\&=\int_{-\infty}^{z}\left[\int_{-\infty}^{\infty}f_\xi(u-y)f_\eta(y)\mathrm{d}y\right]\mathrm{d}u.\end{aligned}$$

由此, 得

$$f_\zeta(z)=\int_{-\infty}^{\infty}f_\xi(z-y)f_\eta(y)\mathrm{d}y, \tag{32}$$

同理, 有

$$f_\zeta(z)=\int_{-\infty}^{\infty}f_\eta(z-x)f_\xi(x)\mathrm{d}x. \tag{33}$$

下面举几个应用这些公式的例子.

设 $\xi_1,\xi_2,\cdots,\xi_n$ 是独立同分布随机变量序列, 其共同的密度是在 $[-1,1]$ 均匀分布密度:

$$f(x)=\begin{cases}\dfrac{1}{2}, & 若\ |x|\leqslant 1,\\ 0, & 若\ |x|>1.\end{cases}$$

那么, 由 (32) 式, 可见

$$f_{\xi_1+\xi_2}(x)=\begin{cases}\dfrac{2-|x|}{4}, & 若\ |x|\leqslant 2,\\ 0, & 若\ |x|>2,\end{cases}$$

$$f_{\xi_1+\xi_2+\xi_3}(x)=\begin{cases}\dfrac{(3-|x|)^2}{16}, & 若\ 1\leqslant|x|\leqslant 3,\\ \dfrac{3-x^2}{8}, & 若\ 0\leqslant|x|\leqslant 1,\\ 0, & 若\ |x|>3,\end{cases}$$

而由归纳法, 可见对于一般情形, 有

$$f_{\xi_1+\cdots+\xi_n}(x)=\begin{cases}\dfrac{1}{2^n(n-1)!}\displaystyle\sum_{k=0}^{\left[\frac{n+x}{2}\right]}(-1)^k\mathrm{C}_n^k(n+x-2k)^{n-1}, & 若\ |x|\leqslant n,\\ 0, & 若\ |x|>n.\end{cases}$$

现在设 $\xi\sim N(m_1,\sigma_1^2),\eta\sim N(m_2,\sigma_2^2)$. 若记

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},$$

则

$$f_\xi(x)=\frac{1}{\sigma_1}\varphi\left(\frac{x-m_1}{\sigma_1}\right),f_\eta(x)=\frac{1}{\sigma_2}\varphi\left(\frac{x-m_2}{\sigma_2}\right),$$

由 (32) 式, 得

$$f_{\xi+\eta}(x)=\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\varphi\left(\frac{x-(m_1+m_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right).$$

于是, *两个独立正态随机变量之和, 仍然是正态随机变量, 其均值为 m_1+m_2, 而方差为 $\sigma_1^2+\sigma_2^2$.*

设 $\xi_1,\xi_2,\cdots,\xi_n$ 是独立同分布随机变量, 都服从均值为 0, 方差为 1 的正态分布. 那么, 由 (26) 式 (用归纳法) 不难求出

$$f_{\xi_1^2+\cdots+\xi_n^2}(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & 若\ x>0,\\ 0, & 若\ x\leqslant 0.\end{cases}\tag{34}$$

随机变量 $\xi_1^2+\cdots+\xi_n^2$ 通常记作 χ_n^2, 而其概率分布称做自由度为 n 的 χ^2分布 (“卡方” 分布) (对照 §3 表 2 − 3).

如果记 $\chi_n=+\sqrt{\chi_n^2}$, 则由 (28) 和 (34) 式, 可见

$$f_{\chi_n}(x)=\begin{cases}\dfrac{2}{2^{n/2}\Gamma(n/2)}x^{n-1}\mathrm{e}^{-\frac{x^2}{2}}, & 若 x>0,\\ 0, & 若 x\leqslant 0.\end{cases}\tag{35}$$

具有此密度的概率分布, 通常称做自由度为 n 的 χ 分布 (“卡” 分布).

仍设独立随机变量 ξ 和 η 各自的密度为 f_ξ 和 f_η. 那么,

$$F_{\xi\eta}(z)=\iint\limits_{\{(x,y):xy\leqslant z\}}f_\xi(x)f_\eta(y)\mathrm{d}x\mathrm{d}y,$$

$$F_{\xi/\eta}(z)=\iint\limits_{\{(x,y):x/y\leqslant z\}}f_\xi(x)f_\eta(y)\mathrm{d}x\mathrm{d}y.$$

由此, 不难得到

$$f_{\xi\eta}(z)=\int_{-\infty}^{\infty} f_{\xi}\left(\frac{z}{y}\right) f_{\eta}(y)\frac{\mathrm{d}y}{|y|}=\int_{-\infty}^{\infty} f_{\eta}\left(\frac{z}{x}\right) f_{\xi}(x)\frac{\mathrm{d}x}{|x|}, \tag{36}$$

和

$$f_{\xi/\eta}(z)=\int_{-\infty}^{\infty} f_{\xi}(zy)f_{\eta}(y)|y|\mathrm{d}y. \tag{37}$$

在 (37) 式中, 设

$$\xi=\xi_0,\quad \eta=\sqrt{\frac{\xi_1^2+\cdots+\xi_n^2}{n}},$$

其中 $\xi_0,\xi_1,\cdots,\xi_n$ 是均值为 0、方差为 $\sigma^2>0$ 的独立正态随机变量. 由 (35) 式, 可得

$$f_t(x)=\frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}. \tag{38}$$

其中随机变量

$$t=\frac{\xi_0}{\sqrt{\dfrac{1}{n}(\xi_1^2+\cdots+\xi_n^2)}}$$

习惯上用 t 表示, 相应的分布称做 t 分布或 "学生" (Student) 分布[①] (对照 §3 表 2 − 3). 注意, 该分布不依赖于 σ.

5. 练习题

1. 验证公式 (9), (10), (24), (27), (28), (34) ~ (38) 的正确性.

2. 设 $\xi_1,\xi_2,\cdots,\xi_n, n\geqslant 2$ 是独立同分布随机变量, 其共同的分布函数为 $F(x)$ (假如密度存在, 则记作 $f(x)$). 记 $\overline{\xi}=\max\{\xi_1,\cdots,\xi_n\}, \underline{\xi}=\min\{\xi_1,\cdots,\xi_n\}, \rho=\overline{\xi}-\underline{\xi}$. 证明

$$F_{\overline{\xi},\underline{\xi}}(y,x)=\begin{cases}(F(y))^n-[F(y)-F(x)]^n, & 若\ y>x,\\ (F(y))^n, & 若\ y\leqslant x,\end{cases}$$

$$f_{\overline{\xi},\underline{\xi}}(y,x)=\begin{cases}n(n-1)[F(y)-F(x)]^{n-2}f(x)f(y), & 若\ y>x,\\ 0, & 若\ y\leqslant x,\end{cases}$$

$$F_{\rho}(x)=\begin{cases}n\displaystyle\int_{-\infty}^{\infty}[F(y)-F(y-x)]^{n-1}f(y)\mathrm{d}y, & 若\ x\geqslant 0,\\ 0, & 若\ x<0,\end{cases}$$

$$f_{\rho}(x)=\begin{cases}n(n-1)\displaystyle\int_{-\infty}^{\infty}[F(y)-F(y-x)]^{n-2}f(y-x)f(y)\mathrm{d}y, & 若\ x>0,\\ 0, & 若\ x\leqslant 0.\end{cases}$$

---

① 哥塞特 (W. S. Gosset, 1876—1937) 是最早创立 t 分布的英国统计学家、化学家、推断统计学派的先驱者, 发表有关论文时使用的笔名是 Student, 因此 t 分布亦称 "学生分布". 哥塞特于 1908 年《生物统计学》(Biometrika, 亦译《生物计量学》) 上发表的论文 "平均值的可能误差" 中首先提出了 t 分布, 并且描绘了其性质. —— 译者

3. 设 ξ_1 和 ξ_2 是独立的泊松随机变量, 其参数相应为 λ_1 和 λ_2. 证明 $\xi_1+\xi_2$ 仍然服从泊松分布, 参数为 $\lambda_1+\lambda_2$.

4. 在 (4) 式中设 $m_1=m_2=0$. 证明

$$f_{\xi/\eta}(z)=\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{\pi(\sigma_2^2z^2-2\rho\sigma_1\sigma_2z+\sigma_1^2)}.$$

5. 量 $\rho^*(\xi,\eta)=\sup_{u,v}\rho(u(\xi),v(\eta))$, 其中对使相关系数 $\rho(u(\xi),v(\eta))$ 有定义的一切博雷尔函数 $u=u(x),v=v(x)$ 求上确界 (sup). $\rho^*(\xi,\eta)$ 称做 ξ 和 η 的**最大相关系数**. 证明随机变量 ξ 和 η 独立, 当且仅当 $\rho^*(\xi,\eta)=0$.

6. 设 $\tau_1,\tau_2,\cdots,\tau_n$ 是独立同分布非负随机变量, 都服从参数为 λ 的指数分布, 其共同的密度为

$$f(t)=\lambda \mathrm{e}^{-\lambda t},\quad t\geqslant 0.$$

证明, 随机变量 $\tau_1+\tau_2+\cdots+\tau_k$ 的分布有密度:

$$\frac{\lambda^k t^{k-1}\mathrm{e}^{-\lambda t}}{(k-1)!},\quad t\geqslant 0,$$

并且

$$\mathbf{P}\{\tau_1+\tau_2+\cdots+\tau_k>t\}=\sum_{i=0}^{k-1}\mathrm{e}^{-\lambda t}\frac{(\lambda t)^i}{i!}.$$

7. 设 $\xi\sim N(0,\sigma^2)$. 证明对于任意 $p\geqslant 1$,

$$\mathbf{E}|\xi|^p=\frac{2^{p/2}}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right)\sigma^p,$$

而

$$\Gamma(s)=\int_0^\infty x^{s-1}\mathrm{e}^{-x}\mathrm{d}x$$

是欧拉 (L. Eular) Γ 函数. 特别, 对于任意 $n\geqslant 1$,

$$\mathbf{E}\xi^{2n}=(2n-1)!!\sigma^{2n}.$$

8. 假设 ξ 和 η 是独立随机变量, 并且 $\xi+\eta$ 的分布与 ξ 的分布是同一分布. 证明 $\eta=0$ (a.c.).

9. 假设 (X,Y) 在单位圆 $\{(x,y):x^2+y^2\leqslant 1\}$ 上有均匀分布, 记 $W=X^2+Y^2$,

$$U=X\sqrt{-\frac{2\ln W}{W}},\quad V=Y\sqrt{-\frac{2\ln W}{W}}.$$

证明, U 和 V 是独立 $N(0,1)-$ 分布随机变量.

10. 设 U 和 V 是独立同在 (0,1) 均匀分布的随机变量. 定义

$$X=\sqrt{-\ln V}\cos(2\pi U),\quad Y=\sqrt{-\ln V}\sin(2\pi U).$$

证明, X 和 Y 是独立 $N(0,1)-$ 分布随机变量.

11. 举一例子: ξ 和 η 是正态随机变量, 但是 $\xi+\eta$ 不服从正态分布.

12. 设 $X_1,\cdots,X_n$ 是独立同分布随机变量, 其共同的密度和分布函数相应为 $f=f(x)$ 和 $F=F(x)$, 而

$$\mathscr{R}_n=\max\{X_1,\cdots,X_n\}-\min\{X_1,\cdots,X_n\}$$

是样本 $(X_1,\cdots,X_n)$ 的 "极差". 证明随机变量 $\mathscr{R}_n$ 的密度 $f_{\mathscr{R}_n}(x), x>0$, 为

$$f_{\mathscr{R}_n}(x)=n(n-1)\int_{-\infty}^{\infty}[F(y)-F(y-x)]^{n-2}f(y)f(y-x)\mathrm{d}y,$$

特别, 对于在 [0,1] 上均匀分布的随机变量 $X_1,\cdots,X_n$,

$$f_{\mathscr{R}_n}(x)=\begin{cases}n(n-1)x^{n-2}(1-x), & 若\ 0\leqslant x\leqslant 1,\\ 0, & 若\ x<0\ 或\ x>1.\end{cases}$$

13. 设 $F(x)$ 是分布函数. 证明, 对于任意 $a>0$, 下列函数也是分布函数:

$$G_1(x)=\frac{1}{a}\int_x^{x+a}F(u)\mathrm{d}u,\quad G_2(x)=\frac{1}{2a}\int_{x-a}^{x+a}F(u)\mathrm{d}u.$$

14. 设随机变量 X 服从参数为 $\lambda>0$ 的指数分布 ($f_X(x)=\lambda\mathrm{e}^{-\lambda x}, x\geqslant 0$), 求随机变量 $Y=X^{1/\alpha}, \alpha>0$ 的概率密度 (相应的分布称做**韦布尔** [W. Weibull] **分布**).

设 $\lambda=1$. 求随机变量 $Y=\ln X$ 的分布密度 (相应的分布称做**双指数分布**).

15. 设随机变量 X 和 Y 的联合分布密度 $f(x,y)$ 形如 $f(x,y)=g(\sqrt{x^2+y^2})$.

(a) 求随机变量 $\rho=\sqrt{X^2+Y^2}$ 和 $\theta=\tan^{-1}(Y/X)$ 的联合分布密度, 证明 ρ 和 θ 独立.

(b) 设 $U=X\cos\alpha+Y\sin\alpha$ 和 $V=-X\sin\alpha+Y\cos\alpha$. 证明随机变量 U 和 V 的联合分布密度与 $f(x,y)$ 相同. (这反映 "随机变量 X 和 Y 的联合分布关于 '旋转' 的不变性" 这一事实.)

16. 设 $X_1,\cdots,X_n$ 是独立同分布随机变量, 其共同的密度和分布函数相应为 $f=f(x)$ 和 $F=F(x)$, 记 (对照练习题 12) $X_{(1)}=\min\{X_1,\cdots,X_n\}$ 是 $X_1,\cdots,X_n$ 的最小值, $X_{(2)}$ 是第二小值, 等等, $X_{(n)}=\max\{X_1,\cdots,X_n\}$ 是 $X_1,\cdots,X_n$ 的最大值 (这样定义的随机变量 $X_{(1)},\cdots,X_{(n)}$ 称做随机变量 $X_1,\cdots,X_n$ 的**顺序统计量**). 证明:

(a) 随机变量 $X_{(k)}$ 的概率分布密度为

$$nf(x)\mathrm{C}_{n-1}^{k-1}[F(x)]^{k-1}[1-F(x)]^{n-k};$$

(b) 随机变量 $X_{(1)},\cdots,X_{(n)}$ 的联合密度为

$$f(x_1,\cdots,x_n)=\begin{cases}n!f(x_1)\cdots f(x_n), & 若\ x_1<\cdots<x_n,\\ 0, & 若不然.\end{cases}$$

17. 设 $X_1, \cdots, X_n$ 是独立同分布随机变量, 且都服从正态分布 $N(\mu, \sigma^2)$. 设

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad \text{其中 } n > 1, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

则随机变量 $\overline{X}$ 和 S^2 分别称做样本均值和样本方差, 证明:

(a) $\mathbf{E}S^2 = \sigma^2$;

(b) $\overline{X}$ 和 S^2 独立;

(c) $\overline{X} \sim N(\mu, \sigma^2/n)$, 而 $(n-1)S^2/\sigma^2$ 服从自由度为 $n-1$ 的 χ^2 分布.

18. 设 $X_1, \cdots, X_n$ 是独立同分布随机变量, $N(N = 1, 2, \cdots)$ 是不依赖于 $X_1, \cdots, X_n$ 的随机变量, 且 $\mathbf{E}N < \infty, \mathbf{D}N < \infty; S_N = X_1 + \cdots + X_N$. 证明

$$\mathbf{D}S_N = \mathbf{D}X_1\mathbf{E}N + (\mathbf{E}X_1)^2\mathbf{D}N, \quad \frac{\mathbf{D}S_N}{\mathbf{E}S_N} = \frac{\mathbf{D}X_1}{\mathbf{E}X_1} + \mathbf{E}X_1\frac{\mathbf{D}N}{\mathbf{E}N}.$$

19. 设 $M(t) = \mathbf{E}e^{tX}$ 是随机变量 X 的母函数. 证明对于任意 $t > 0$, 有 $\mathbf{P}\{X \geqslant 0\} \leqslant M(t)$.

20. 设 $X, X_1, \cdots, X_n$ 是独立同分布随机变量, $S_n = X_1 + \cdots + X_n, S_0 = 0$, 而

$$\overline{M}_n = \max_{0 \leqslant j \leqslant n} S_j, \quad \overline{M} = \sup_{n \geqslant 0} S_n.$$

证明

(a) 对于 $n \geqslant 1, \overline{M}_n$ 和 $(\overline{M}_{n-1} + X)^+$ 同分布;

(b) 如果 $S_n \to \infty(\mathbf{P} - \text{a.c.})$, 则 $\overline{M}$ 和 $(\overline{M} + X)^+$ 同分布;

(c) 如果 $-\infty < \mathbf{E}X < 0$ 且 $\mathbf{E}X^2 < \infty$, 则

$$\mathbf{E}\overline{M} = \frac{\mathbf{D}X - \mathbf{D}(S + X)^-}{-2\mathbf{E}X}.$$

21. 在上题的条件下, 对于任意 $\varepsilon > 0$, 设

$$\overline{M}(\varepsilon) = \sup_{n \geqslant 0}(S_n - n\varepsilon).$$

证明

$$\lim_{\varepsilon \downarrow 0} \varepsilon\overline{M}(\varepsilon) = \frac{\mathbf{D}X}{2}.$$

§9. 建立具有给定有限维分布的过程

1. 具有给定分布函数的随机变量存在性 设 $\xi = \xi(\omega)$ 是概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的随机变量. 而

$$F_\xi(x) = \mathbf{P}\{\omega : \xi(\omega) \leqslant x\}$$

是其分布函数. 自然, 在 §3 定义 1 的意义上, $F_\xi(x)$ 是数轴上的分布函数.

现在提出如下问题. 假设 $F = F(x)$ 是 $\mathbb{R}$ 上某一分布函数. 问是否存在以 $F(x)$ 为分布函数的随机变量?

证实问题的这一提法的一个理由如下. 许多概率论的命题从如下的句子开始: "*设 ξ 是分布函数为 $F(x)$ 的随机变量, 则 ……*". 因此, 为使类似的论断有意义, 应该确信所考虑的对象确实存在. 由于为给出随机变量, 应先给出其定义域 $(\Omega, \mathscr{F})$; 而为了提起其概率分布, 应该有 $(\Omega, \mathscr{F})$ 上的概率测度 $\mathbf{P}$. 于是, 关于具有给定分布函数 $F(x)$ 的随机变量存在性的正确提法是:

是否存在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 和此概率空间上的随机变量 $\xi = \xi(\omega)$, 使

$$\mathbf{P}\{\omega : \xi(\omega) \leqslant x\} = F(x)?$$

现在证明, 对于所提问题的答案是肯定的, 其实答案已经包含在 §3 的定理 1 中. 事实上, 设

$$\Omega = \mathbb{R}, \quad \mathscr{F} = \mathscr{B}(\mathbb{R}),$$

那么, 由 §3 的定理 1 可见, 在 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上存在 (并且唯一) 概率测度 $\mathbf{P}$, 使对于任意 $a < b, \mathbf{P}(a, b] = F(b) - F(a)$.

设 $\xi(\omega) = \omega$. 那么,

$$\mathbf{P}\{\omega : \xi(\omega) \leqslant x\} = \mathbf{P}\{\omega : \omega \leqslant x\} = \mathbf{P}(-\infty, x] = F(x).$$

于是, 建立了所要求的概率空间和随机变量.

2. 具有给定有限维分布的随机过程的存在性 现在对于随机过程提出类似的问题.

对于 $t \in T \subseteq \mathbb{R}$, 设 $X = (\xi_t)_{t \in T}$ 是概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的随机过程 (见 §5 定义 3).

按照物理的观点, 随机过程最重要的概率特征是其*有限维分布函数族* $\{F_{t_1, \cdots, t_n}(x_1, \cdots, x_n)\}$:

$$F_{t_1, \cdots, t_n}(x_1, \cdots, x_n) = \mathbf{P}\{\omega : \xi_{t_1} \leqslant x_1, \cdots, \xi_{t_n} \leqslant x_n\}, \tag{1}$$

其中对于一切数组 $t_1, \cdots, t_n, t_1 < \cdots < t_n, F_{t_1, \cdots, t_n}(x_1, \cdots, x_n)$ 是给定的.

由 (1) 式可见, 对于一切数组 $t_1, \cdots, t_n (t_1 < \cdots < t_n), F_{t_1, \cdots, t_n}(x_1, \cdots, x_n)$ 是 n 元分布函数 (见 §3 定义 2), 而且分布函数族 $\{F_{t_1, \cdots, t_n}(x_1, \cdots, x_n)\}$ 满足如下一致性条件 (见 §3 (20) 式):

$$\begin{aligned} &F_{t_1, \cdots, t_k, \cdots, t_n}(x_1, \cdots, \infty, \cdots, x_n) \\ &= F_{t_1, \cdots, t_{k-1}, t_{k+1}, \cdots, t_n}(x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_n). \end{aligned} \tag{2}$$

现在, 自然地提出这样的问题; 分布函数 $F_{t_1, \cdots, t_n}(x_1, \cdots, x_n)$ 族 $\{F_{t_1, \cdots, t_n}(x_1, \cdots, x_n)\}$ (见 §3 定义 2), 在什么条件下可以做某一随机过程的有限维分布函数族? 非常出色的是, 一致性条件 (2) 可以穷尽一切附加条件.

定理 1 (柯尔莫戈洛夫关于过程存在定理性) 设 $\{F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)\}$ 是给定的满足一致性条件 (2) 的有限维分布函数族, 其中 $t_i \in T \subseteq \mathbb{R}, n \geqslant 1, t_1 < t_2 < \cdots < t_n$. 那么, 在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上存在随机过程 $X = (\xi_t)_{t\in T}$, 使

$$\mathbf{P}\{\omega : \xi_{t_1} \leqslant x_1, \cdots, \xi_{t_n} \leqslant x_n\} = F_{t_1,\cdots,t_n}(x_1,\cdots,x_n). \tag{3}$$

证明 设

$$\Omega = \mathbb{R}^T, \quad \mathscr{F} = \mathscr{B}(\mathbb{R}^T).$$

即作为空间 Ω 取实函数的空间 $\omega = (\omega_t)_{t\in T}$, 而 $\mathscr{F}$ 取由柱集生成的 $\sigma-$ 代数.

设 $\tau = [t_1, t_2, \cdots, t_n], t_1 < t_2 < \cdots < t_n$. 那么, 根据 §3 定理 2. 在空间 $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ 中可以建立 (并且是唯一的) 概率测度 P_τ, 使

$$P_\tau\{(\omega_{t_1},\cdots,\omega_{t_n}) : \omega_{t_1} \leqslant x_1, \cdots, \omega_{t_n} \leqslant x_n\} = F_{t_1,\cdots,t_n}(x_1,\cdots,x_n). \tag{4}$$

由一致性条件 (2) 可见, 概率测度 $\{P_\tau\}$ 族也满足一致性条件 (见 §3 的 (20) 式). 根据 §3 的定理 4, 在空间 $(\mathbb{R}^T, \mathscr{B}(\mathbb{R}^T))$ 上存在概率测度 $\mathbf{P}$, 使对于 $\tau = [t_1,\cdots,t_n], t_1 < \cdots < t_n$,

$$\mathbf{P}\{\omega : (\omega_{t_1},\cdots,\omega_{t_n}) \in B\} = P_\tau(B).$$

由此亦可见, 条件 (4) 成立. 因此, 作为欲求的随机过程 $X = (\xi_t(\omega))_{t\in T}$, 可以取如下的随机过程:

$$\xi_t(\omega) = \omega_t, \quad t \in T. \tag{5}$$

□

注 1 所建立的概率空间 $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \mathbf{P})$ 常称做*标准的*, 而由 (5) 式表示随机过程方法常称为建立过程的*坐标方法*.

注 2 设 $(E_\alpha, \mathscr{E}_\alpha)$ 是完全可分度量空间, 而 α 属于任意下标的集合 $\mathfrak{U}$. 设 $\{P_\tau\}$ 是 $(E_{\alpha_1} \times \cdots \times E_{\alpha_n}, \mathscr{E}_{\alpha_1} \otimes \cdots \otimes \mathscr{E}_{\alpha_n})$ 上的有限维分布函数族: $P_\tau, \tau = [t_1,\cdots,t_n]$. 那么, 存在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 和 $\mathscr{F}/\mathscr{E}_\alpha-$ 可测函数族 $(X_\alpha(\omega))_{\alpha\in\mathfrak{U}}$, 使对于任意 $\tau = [\alpha_1,\cdots,\alpha_n]$ 和 $B \in \mathscr{E}_{\alpha_1} \otimes \cdots \otimes \mathscr{E}_{\alpha_n}$, 有

$$\mathbf{P}\{(X_{\alpha_1},\cdots,X_{\alpha_n}) \in B\} = P_\tau(B).$$

如果对于每一个 $\omega = (\omega_\alpha), \alpha \in \mathfrak{U}$, 设 $\Omega = \prod\limits_\alpha E_\alpha, \mathscr{F} = \bigotimes\limits_\alpha \mathscr{E}_\alpha$ 和 $X_\alpha(\omega) = \omega_\alpha$, 则这一结果由 §3 的定理 4 得到, 而它推广了定理 1 的命题.

系 1 *设 $F_1(x), F_2(x), \cdots$ 是一元分布函数序列. 那么, 存在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 和独立随机变量序列 $\xi_1, \xi_2, \cdots$, 使*

$$\mathbf{P}\{\omega : \xi_i(\omega) \leqslant x\} = F_i(x). \tag{6}$$

特别, 存在概率空间 $(\Omega,\mathscr{F},\mathbf{P})$, 在 $(\Omega,\mathscr{F},\mathbf{P})$ 上定义了无限伯努利随机变量序列 (参见第一章 §5 第 2 小节). 这里, 作为 Ω 可以取空间:

$$\Omega=\{\omega:\omega=(a_1,a_2,\cdots),a_i=0,1\}$$

(参见定理 2).

为证明该系, 只需设 $F_{1,\cdots,n}(x_1,\cdots,x_n)=F_1(x_1)\cdots F_n(x_n)$, 并且应用定理 1.

系 2 设 $T=[0,\infty)$, 而对于 $s,t\in T,t>s,x\in\mathbb{R},B\in\mathscr{B}(\mathbb{R}).\{P(s,x;t,B)\}$ 是非负函数族, 并且满足条件:

a) 对于固定的 $s,x,t,P(s,x;t,B)$ 是 B 的概率测度;

b) 对于固定的 s,t 和 $B,P(s,x;t,B)$ 是 x 的博雷尔函数;

c) 对于一切 $0\leqslant s<t<\tau$ 和 $B\in\mathscr{B}(\mathbb{R})$, 满足柯尔莫戈洛夫 – 查普曼方程:

$$P(s,x;\tau,B)=\int_{\mathbb{R}}P(s,x;t,\mathrm{d}y)P(t,y;\tau,B). \tag{7}$$

此外, 假设 $\pi=\pi(\cdot)$ 是 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的概率测度.

那么, 存在概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 和随机过程 $X=(\xi_t)_{t\geqslant0}$, 使对于 $0=t_0<t_1<\cdots<t_n$, 有

$$\begin{aligned}&\mathbf{P}\{\xi_{t_0}\leqslant x_0,\xi_{t_1}\leqslant x_1,\cdots,\xi_{t_n}\leqslant x_n\}\\&=\int_{-\infty}^{x_0}\pi(\mathrm{d}y_0)\int_{-\infty}^{x_1}P(0,y_0;t_1,\mathrm{d}y_1)\cdots\int_{-\infty}^{x_n}P(t_{n-1},y_{n-1};t_n,\mathrm{d}y_n).\end{aligned} \tag{8}$$

这样建立的过程 X 称做马尔可夫过程, π 称做其初始分布, 而 $\{P(s,x;t,B)\}$ 称做过程的转移概率族.

系 3 设 $T=\{0,1,2,\cdots\}$, 而对于 $k\geqslant1,x\in\mathbb{R},B\in\mathscr{B}(\mathbb{R}),\{P_k(x;B)\}$ 是非负函数族, 并且 (对于固定的 k 和 x) $P_k(x;B)$ 是 B 的概率测度, 而 (对于固定的 k 和 B) $P_k(x;B)$ 是 x 的可测函数. 此外, 假设 $\pi=\pi(\cdot)$ 是 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的概率测度.

那么, 可以建立概率空间 $(\Omega,\mathscr{F},\mathbf{P})$, 使对于 $(\Omega,\mathscr{F},\mathbf{P})$ 上的随机变量族 $X=(\xi_0,\xi_1,\cdots)$, 有

$$\begin{aligned}&\mathbf{P}\{\xi_0\leqslant x_0,\xi_1\leqslant x_1,\cdots,\xi_n\leqslant x_n\}\\&=\int_{-\infty}^{x_0}\pi(\mathrm{d}y_0)\int_{-\infty}^{x_1}P_1(y_0;\mathrm{d}y_1)\cdots\int_{-\infty}^{x_n}P_n(y_{n-1};\mathrm{d}y_n).\end{aligned}$$

3. 测度的开拓和随机序列的存在性 由系 1 知, 存在分布函数相应为 $F_1,F_2,\cdots$ 的独立随机变量 $\xi_1,\xi_2,\cdots$ 序列.

现在设 $(E_1,\mathscr{E}_1),(E_2,\mathscr{E}_2),\cdots$ 是完全可分度量空间, 而 $P_1,P_2,\cdots$ 是在这些空间上的概率测度. 那么. 由注 2 可见, 存在概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 和独立随机元序列 $X_1,X_2,\cdots$, 使 X_n 为 $\mathscr{F}/\mathscr{E}_n-$ 可测, 且 $\mathbf{P}\{X_n\in B\}=P_n(B),B\in\mathscr{E}_n$. 实际上, 当 $(E_n,\mathscr{E}_n)$ 是任意可测空间时. 这一结果仍然成立.

定理 2 (图尔恰 [I. Tulcea] 关于测度的开拓和随机序列的存在性)　设 $(\Omega_n, \mathscr{F}_n), n=1,2,\cdots$ 是任意可测空间, $\Omega=\prod\Omega_n, \mathscr{F}=\bigotimes\mathscr{F}_n$. 假设在 $(\Omega_1,\mathscr{F}_1)$ 上给定概率测度 P_1, 而对于 $(\omega_1,\cdots,\omega_n)\in\Omega=\Omega_1\times\cdots\times\Omega_n, n\geqslant 1$, 在 $(\Omega_{n+1},\mathscr{F}_{n+1})$ 上给定概率测度 $P(\omega_1,\cdots,\omega_n;\cdot)$; 并且对于每一个 $B\in\mathscr{F}_{n+1}, P(\omega_1,\cdots,\omega_n;B)$ 是 $(\omega_1,\cdots,\omega_n)$ 的可测函数. 对于 $A_i\in\mathscr{F}_i, n\geqslant 1$, 设

$$P_n(A_1\times\cdots\times A_n)=\int_{A_1}P_1(\mathrm{d}\omega_1)\int_{A_2}P(\omega_1;\mathrm{d}\omega_2)\cdots\int_{A_n}P(\omega_1,\cdots,\omega_{n-1};\mathrm{d}\omega_n). \quad (9)$$

那么, 在 $(\Omega,\mathscr{F})$ 上存在唯一概率测度 $\mathbf{P}$, 使对于任意 $n\geqslant 1$,

$$\mathbf{P}\{\omega:\omega_1\in A_1,\cdots,\omega_n\in A_n\}=P_n(A_1\times\cdots\times A_n), \quad (10)$$

并存在随机变量序列 $X=(X_1(\omega),X_2(\omega),\cdots)$, 使对于 $A_i\in\mathscr{E}_i$,

$$\mathbf{P}\{\omega:X_1(\omega)\in A_1,\cdots,X_n(\omega)\in A_n\}=P_n(A_1\times\cdots\times A_n). \quad (11)$$

证明　1) 证明的第一步: 对于每一个 $n>1$, 证明由 (9) 式在 "矩形" $A_1\times\cdots\times A_n$ 上给定的集函数 P_n, 可以开拓到 $\sigma-$ 代数 $\mathscr{F}_1\otimes\cdots\otimes\mathscr{F}_n$ 上.

为此, 对于每一个 $n\geqslant 2$ 和 $B\in\mathscr{F}_1\otimes\cdots\otimes\mathscr{F}_n$, 设

$$\begin{aligned}P_n(B)=&\int_{\Omega_1}P_1(\mathrm{d}\omega_1)\int_{\Omega_2}P(\omega_1;\mathrm{d}\omega_2)\cdots\int_{\Omega_{n-1}}P(\omega_1,\cdots,\omega_{n-2};\mathrm{d}\omega_{n-1})\\&\times\int_{\Omega_n}I_B(\omega_1,\cdots,\omega_{n-1},\omega_n)P(\omega_1,\cdots,\omega_{n-1};\mathrm{d}\omega_n). \end{aligned}\quad (12)$$

易见, 对于 $B=A_1\times\cdots\times A_n$, (12) 式的右侧与 (9) 式的右侧相同. 此外. 对于 $n=2$. 如同 §6 的定理 8, P_2 是测度. 于是, 利用归纳法容易证明, 对于每一个 $n\geqslant 2$, P_n 是测度.

2) 证明的第二步: 与关于在 $(\mathbb{R}^\infty,\mathscr{B}(\mathbb{R}^\infty))$ 中测度开拓的柯尔莫戈洛夫定理 (§3 定理 3) 相同. 具体地说, 对于任意柱集 $J_n(B)=\{\omega\in\Omega:(\omega_1,\cdots,\omega_n)\in B\}, B\in\mathscr{F}_1\otimes\cdots\otimes\mathscr{F}_n$, 利用等式

$$\mathbf{P}(J_n(B))=P_n(B) \quad (13)$$

定义集函数 $\mathbf{P}$. 由 (12) 式, 以及 $P(\omega_1,\cdots,\omega_k;\cdot)$ 是测度, 容易证明定义 (13) 具有一致性: $\mathbf{P}(J_n(B))$ 的值不依赖于 (13) 式中柱集的表示方法.

由 (13) 式在柱集上定义的集函数 $\mathbf{P}$, 显然也是定义在包含所有柱集的代数上的集函数 $\mathbf{P}$. 由此可见, 集函数 $\mathbf{P}$ 在此代数上是有限可加测度. 只剩下验证集函数 $\mathbf{P}$ 在此代数上的可数可加性, 然后应用卡拉泰奥多里定理.

在 §3 的定理 3 进行过上面提到的证明; 证明基于空间 $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 的如下性质: 对于每一个博雷尔集 B, 存在测度任意接近 B 的测度的紧统 $A\subseteq B$. 对于现在的情形. 这一方法有如下形式变化.

像 §3 的定理 3 一样, 设 $\{\widehat{B}_n\}_{n\geqslant 1}$ 是递减为空集 $\varnothing$ 的柱集序列,

$$\widehat{B}_n = \{\omega : (\omega_1, \cdots, \omega_n) \in B_n\}$$

但是

$$\lim_{n\to\infty} \mathbf{P}(\widehat{B}_n) > 0. \tag{14}$$

由 (12) 式可见, 对于 $n > 1$, 有

$$\mathbf{P}(\widehat{B}_n) = \int_{\Omega_1} f_n^{(1)}(\omega_1) P_1(\mathrm{d}\omega_1).$$

其中

$$f_n^{(1)}(\omega_1) = \int_{\Omega_2} P(\omega_1; \mathrm{d}\omega_2) \cdots \int_{\Omega_n} I_{B_n}(\omega_1, \cdots, \omega_{n-1}, \omega_n) P(\omega_1, \cdots, \omega_{n-1}; \mathrm{d}\omega_n).$$

由于 $\widehat{B}_{n+1} \subseteq \widehat{B}_n$. 则 $B_{n+1} \subseteq B_n \times \Omega_{n+1}$. 因而 $I_{B_{n+1}}(\omega_1, \cdots, \omega_{n+1}) \leqslant I_{B_n}(\omega_1, \cdots, \omega_n) I_{\Omega_{n+1}}(\omega_{n+1})$. 因此. 函数序列 $\{f_n^{(1)}(\omega_1)\}_{n\geqslant 1}$ 是减小的. 设

$$f^{(1)}(\omega_1) = \lim_n f_n^{(1)}(\omega_1).$$

那么, 根据控制收敛定理

$$\lim_n \mathbf{P}(\widehat{B}_n) = \lim_n \int_{\Omega_1} f_n^{(1)}(\omega_1) P_1(\mathrm{d}\omega_1) = \int_{\Omega_1} f^{(1)}(\omega_1) P_1(\mathrm{d}\omega_1).$$

根据假设, $\lim_n \mathbf{P}(\widehat{B}_n) > 0$. 由此可见, 由于若 $\omega_1 \notin B_1$, 则对于所有 $n \geqslant 1$, $f_n^{(1)}(\omega_1) = 0$, 故存在 $\omega_1^0 \in B_1$, 使 $f^{(1)}(\omega_1^0) > 0$.

其次, 对于 $n > 2$,

$$f_n^{(1)}(\omega_1^0) = \int_{\Omega_2} f_n^{(2)}(\omega_2) P(\omega_1^0; \mathrm{d}\omega_2), \tag{15}$$

其中

$$\begin{aligned} f_n^{(2)}(\omega_2) &= \int_{\Omega_3} P(\omega_1^0, \omega_2; \mathrm{d}\omega_3) \cdots \\ &\quad \cdots \int_{\Omega_n} I_{B_n}(\omega_1^0, \omega_2, \cdots, \omega_n) P(\omega_1^0, \omega_2, \cdots, \omega_{n-1}; \mathrm{d}\omega_n). \end{aligned}$$

同序列的 $\{f_n^{(1)}(\omega_1)\}$ 的情形一样, 证明序列 $\{f_n^{(2)}(\omega_2)\}$ 是减少的. 设

$$f^{(2)}(\omega_2) = \lim_{n\to\infty} f_n^{(2)}(\omega_2).$$

那么. 由 (15) 式, 可见

$$0 < f^{(1)}(\omega_1^0) = \int_{\Omega_2} f^{(2)}(\omega_2) P(\omega_1^0; \mathrm{d}\omega_2),$$

且存在点 $\omega_2^0 \in \Omega_2$, 使 $f^{(2)}(\omega_2^0) > 0$. 这时 $(\omega_1^0, \omega_2^0) \in B_2$. 继续这一过程, 对于任意 $n \geqslant 1$, 存在点 $(\omega_1^0, \cdots, \omega_n^0) \in B_n$. 从而, 点 $(\omega_1^0, \cdots, \omega_n^0, \cdots) \in \bigcap \widehat{B}_n$. 但是, 根据假设 $\bigcap \widehat{B}_n = \varnothing$. 出现的矛盾说明 $\lim_n \mathbf{P}(\widehat{B}_n) = 0$.

于是, 定理涉及概率测度 $\mathbf{P}$ 存在性的部分得证.

3) 定理的结尾部分显然可以由上面的结果得到. 为此只需要设 $X_n(\omega) = \omega_n$, $n \geqslant 1$. □

系 4 设 $(E_n, \mathscr{E}_n)_{n \geqslant 1}$ 是任意可测空间, $(P_n)_{n \geqslant 1}$ 是该空间上的概率测度. 那么, 存在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 和相应地取值于可测空间 $(E_1, \mathscr{E}_1), (E_2, \mathscr{E}_2), \cdots$ 的独立随机元族 $X_1, X_2, \cdots$, 并且

$$\mathbf{P}\{\omega : X_n(\omega) \in B\} = P_n(B), \quad B \in \mathscr{E}_n, \ n \geqslant 1.$$

系 5 设 $E = \{1, 2, \cdots\}, \{p_k(x, y)\}, k \geqslant 1, x, y \in E$ 是非负函数族, 并且满足

$$\sum_{y \in E} p_k(x, y) = 1, x \in E, k \geqslant 1.$$

此外, 设 $\pi = \pi(\cdot)$ 是 E 上的概率分布 $(\pi(x) \geqslant 0, \sum_{x \in E} \pi(x) = 1)$.

那么, 存在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$, 且在此空间上存在随机变量族 $X = \{\xi_0, \xi_1, \cdots\}$, 使对于一切 $x_i \in E, n \geqslant 1$, 有

$$\mathbf{P}\{\xi_0 = x_0, \xi_1 = x_1, \cdots, \xi_n = x_n\} = \pi(x_0) p_1(x_0, x_1) \cdots p_n(x_{n-1}, x_n) \tag{16}$$

(对照第一章 §12 之 (4) 式). 作为空间 Ω 可以取

$$\Omega = \{\omega : \omega = (x_0, x_1, \cdots), x_i \in E\}.$$

满足 (16) 式的随机变量序列 $X = \{\xi_0, \xi_1, \cdots\}$, 称做具有可数状态集 E 的马尔可夫链, $\{p_k(x, y)\}$ 称为其转移概率矩阵, 而称为其初始概率分布 π. (对照第一章 §12 的定义, 以及第八章 §1 的定义.)

4. 更新过程 柯尔莫戈洛夫定理 (定理 1) 指出了, 具有给定有限维分布函数的过程的存在性. 这时, 定理的证明用到典则概率空间, 而构造过程使用坐标式方法. 这本身就说明了过程构造的复杂性.

按这种观点, 在尽量少地使用 "概率结构" 的情况下, 构建具有给定性质的随机过程的情形会引起人们极大兴趣.

为演示这种可能性, 我们现在考虑所谓*更新过程*. (其特殊情形是泊松过程; 见第二章 §10.)

设 $(\sigma_1, \sigma_2, \cdots)$ 是独立同分布随机变量序列, 其共同的分布函数为 $F = F(x)$. (定理 1 的系 1 保障这样随机变量序列的存在.)

由 $(\sigma_1, \sigma_2, \cdots)$ 序列形成新的序列 $(T_0, T_1, \cdots)$, 其中 $T_0 = 0$, 且

$$T_n = \sigma_1 + \cdots + \sigma_n, n \geqslant 1.$$

为直观计, 我们把 T 视为时间: T_n 表示第 n 个质点出现的时间 (例如, 电话的第 n 次呼唤). 那么, σ_n 描绘第 $(n-1)$ 到第 n 个质点 (第 n 次呼唤) 持续的时间.

习惯上把随机过程 $N=(N_t)_{t\geqslant 0}$, 其中 N_t 是 (构造性地给定的) 量

$$N_t=\sum_{n=1}^{\infty} I(T_n\leqslant t), \tag{17}$$

称做*更新过程*.

显然, N_t 亦可定义为

$$N_t=\max\{n: T_n\leqslant t\}, \tag{18}$$

即 N_t 是出现在区间 $(0,t]$ 上的质点 (呼唤) 数; 这时显然

$$\{N_t\geqslant n\}=\{T_n\leqslant t\}. \tag{19}$$

这一简单的公式十分有用, 因为它可以把对随机过程 $N=(N_t)_{t\geqslant 0}$ 概率性质的研究, 归结为对独立随机变量 $\sigma_1,\cdots,\sigma_n (n\geqslant 1)$ 之和 $T_n=\sigma_1+\cdots+\sigma_n$ 的性质的研究 (参见第四章 §3 第 4 小节, 以及第七章 §2 第 4 小节).

由 (17) 式立即可见, *更新函数* $m(t)=\mathbf{E}N_t (t\geqslant 0)$ 决定于分布函数 $F_n(t)=\mathbf{P}\{T_n\leqslant t\}$ 如下:

$$m(t)=\sum_{n=1}^{\infty} F_n(t). \tag{20}$$

5. 练习题

1. 设 $\Omega=[0,1]$.$\mathscr{F}$ 是 $[0,1]$ 上的博雷尔集类, $\mathbf{P}$ 是 $[0,1]$ 上的勒贝格测度. 我们称空间 $(\Omega,\mathscr{F},\mathbf{P})$ 为 "通用的", 如果对于 $(\Omega,\mathscr{F},\mathbf{P})$ 上的任意分布函数 $F(x)$, 可以定义一随机变量 $\xi=\xi(\omega)$, 使其分布函数 $F_\xi(x)=\mathbf{P}\{\xi\leqslant x\}$ 恰好是 $F(x)$. 证明空间 $(\Omega,\mathscr{F},\mathbf{P})$ 为 "通用的". (*提示*: $\xi(\omega)=F^{-1}(\omega)$, 其中 $F^{-1}(\omega)=\sup\{x: F(x)<\omega\}, 0<\omega<1$, 而 $\xi(0),\xi(1)$ 可以是任意的.)

2. 验证定理 1 和 2 的系中分布族的一致性.

3. 由定理 1 导出定理 2 的系 2 的命题.

4. 设 F_n 是随机变量 $T_n (n\geqslant 1)$ 的分布函数 (第 4 小节). 证明

$$F_{n+1}(t)=\int_0^t F_n(t-s)\mathrm{d}F(s), n\geqslant 1, \text{其中 } F_1=F.$$

5. 证明 $\mathbf{P}\{N_t=n\}=F_n(t)-F_{n+1}(t)$ (见 (17) 式).

6. 证明, 在上面第 4 小节引进的更新函数 $m(t)$ 满足*更新方程*:

$$m(t)=F(t)+\int_0^t m(t-x)\mathrm{d}F(x). \tag{21}$$

7. 证明, 在有限区间上有界的函数类中, 由 (20) 式定义的函数是方程 (21) 的唯一解.

8. 设 T 是任意集合.

(i) 假设对于每个 $t\in T$, 给定一概率空间 $(\Omega_t,\mathscr{F}_t,\mathbf{P}_t)$. 记

$$\Omega=\prod_{t\in T}\Omega_t,\quad \mathscr{F}=\bigotimes_{t\in T}\mathscr{F}_t.$$

证明, 在 $(\Omega,\mathscr{F})$ 上存在唯一概率测度 $\mathbf{P}$, 使

$$\mathbf{P}\left(\prod_{t\in T}B_t\right)=\prod_{t\in T}\mathbf{P}(B_t),$$

其中对于除有限个之外的一切下标 $t, B_t\in\mathscr{F}_t, t\in T, B_t=\Omega_t$. (提示: 这里 $\mathbf{P}$ 在相应的代数上, 并利用定理 2 的证明方法.)

(ii) 假设对于每个 $t\in T$, 给定可测空间 $(E_t,\mathscr{E}_t)$ 和该空间上的概率测度 $\mathbf{P}_t$. 证明存在概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 及独立随机元素 $(X_t)_{t\in T}$, 使 X_t 为 $\mathscr{F}_t/\mathscr{E}_t-$ 可测的, 且 $\mathbf{P}\{X_t\in B\}=\mathbf{P}_t(B), B\in\mathscr{E}_t$.

§10. 随机变量序列收敛的各种形式

1. 收敛性的基本类型　同数学分析一样, 在概率论中也需要考虑随机变量不同形式的收敛性. 现在, 讨论如下形式的收敛性: *依概率收敛*, *依概率 1 收敛*, *p 阶平均收敛*, *按分布收敛*.

从定义开始叙述. 设 $\xi,\xi_1,\xi_2,\cdots$ 是某一概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 上的随机变量.

定义 1　随机变量序列 $\xi_1,\xi_2,\cdots$ (亦记作 $\{\xi_n\},\{\xi_n\}_{n\geqslant1}$ 或 $(\xi_n),(\xi_n)_{n\geqslant1}$), 称做*依概率收敛*于随机变量 ξ 的 (记作 $\xi_n\xrightarrow{\mathbf{P}}\xi$), 如果对于任意 $\varepsilon>0$, 有

$$\mathbf{P}\{|\xi_n-\xi|>\varepsilon\}\to0,\quad n\to\infty. \tag{1}$$

我们已经遇到过这种形式的收敛性, 即伯努利大数定律:

$$\mathbf{P}\left\{\left|\frac{S_n}{n}-p\right|>\varepsilon\right\}\to0,\quad n\to\infty$$

(记号见第一章 §5). 在数学分析中这种形式的收敛性, 称做*依测度收敛*.

定义 2　随机变量序列 $\xi_1,\xi_2,\cdots$, 称做*依概率 1 (几乎必然或几乎处处) 收敛*于随机变量 ξ, 如果

$$\mathbf{P}\{\omega:\xi_n\nrightarrow\xi\}\to0,\quad n\to\infty, \tag{2}$$

即如果使 $\xi_n(\omega)$ 收敛于 $\xi(\omega)$ 的结局 ω 的集合的概率等于 1.

这种形式 (依概率 1) 的收敛性, 有多种记号:

$$\xi_n \to \xi(\mathbf{P}-\text{a.c.}),\ 或\ \xi_n \to \xi(\text{a.c.}),\ 或\ \xi_n \xrightarrow{\text{a.c.}} \xi,\ 或\ \xi_n \xrightarrow{\text{a.e.}} \xi.$$

定义 3 随机变量序列 $\xi_1, \xi_2, \cdots$, 称做 p 阶平均收敛于随机变量 ξ, 如果

$$\mathbf{E}|\xi_n - \xi|^p \to 0, \quad n \to \infty. \tag{3}$$

在数学分析中这种形式的收敛性称做 L^p- 收敛. 因此, (3) 式可写为 $\xi_n \xrightarrow{L^p} \xi$. 特别, 若 $p=2$, 则这种收敛称做平方平均收敛 (或均方收敛), 记作 $\xi = \text{l.i.m.}\xi_n$ (l.i.m. 是 "limit in mean" —— 平均极限的缩写).

定义 4 随机变量序列 $\xi_1, \xi_2, \cdots$, 称按分布 (或依分布) 收敛于随机变量 ξ, 如果对于任意有界连续函数 $f = f(x)$, 有

$$\mathbf{E}f(\xi_n) \to \mathbf{E}f(\xi), \quad n \to \infty \tag{4}$$

记作: $\xi_n \xrightarrow{d} \xi, \xi_n \xrightarrow{law} \xi$ 其中 d 是 distribution (分布) 的字头, law 是 (分布) 律.

这一收敛性名称的来历, 像将由第三章 §1 表明的那样, 条件 (4) 等价于, 分布函数 $F_{\xi_n}(x)$ 在分布函数 $F_\xi(x)$ 的每一个连续点上收敛于 $F_\xi(x)$ (称做 $F_{\xi_n}(x)$ 基本收敛于 $F_\xi(x)$, 记作 $F_{\xi_n} \Rightarrow F_\xi$).

应该强调, 随机的按分布收敛性, 只有通过其分布函数的收敛性定义. 因此, 关于这种形式的收敛性, 当且仅当随机变量定义在不同的概率空间上才有意义. 这种形式的收敛性, 在第三章将详细研究, 到时要特别说明, 为什么在收敛性 $F_{\xi_n} \Rightarrow F_\xi$ 的定义中, 仅要求在 $F_\xi(x)$ 的连续点上收敛, 而不是在一切 x 点上.

2. 基本随机变量序列的概念 在数学分析中, 在解决给定函数序列 (在一定意义下) 收敛性的问题时, 基本序列 (或柯西序列) 的概念是非常重要的. 对于现在所研究的随机变量序列收敛性的情形, 也引进类似的概念.

称随机变量序列 $\{\xi_n\}_{n\geqslant 1}$ (或简记为 $\{\xi_n\}$) 为依概率为基本的, 如果满足条件: 对于任意 $\varepsilon > 0$, 有 $\mathbf{P}\{|\xi_n - \xi_m| \geqslant \varepsilon\} \to 0, n, m \to \infty$; 称 $\{\xi_n\}_{n\geqslant 1}$ 为依概率 1 为基本的, 若对于几乎一切 $\omega \in \Omega, \{\xi_n(\omega)\}_{n\geqslant 1}$ 为基本序列; 称 $\{\xi_n(\omega)\}_{n\geqslant 1}$ 在 L^p (或 $p(0<p<\infty)$ 阶平均) 的意义上为基本的, 若 $\mathbf{E}|\xi_n - \xi_m|^p \to 0, n, m \to \infty, 0 < p < \infty$.

3. 随机变量序列的依概率 1 收敛

定理 1 a) 使 $\xi_n \to \xi(\mathbf{P}-\text{a.c.})$ 的充分和必要条件是, 对于任意 $\varepsilon > 0$,

$$\mathbf{P}\left\{\sup_{k\geqslant n} |\xi_k - \xi| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty. \tag{5}$$

b) 随机变量序列 $\{\xi_n\}_{n\geqslant 1}$ 依概率 1 为基本的, 当且仅当对于任意 $\varepsilon > 0$,

$$\mathbf{P}\left\{\sup_{k\geqslant n, l\geqslant n} |\xi_k - \xi_l| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty, \tag{6}$$

或等价地

$$\mathbf{P}\left\{\sup_{k\geqslant 0}|\xi_{n+k}-\xi_n|\geqslant\varepsilon\right\}\to 0,\quad n\to\infty. \tag{7}$$

证明 a) 设 $A_n^\varepsilon=\{\omega:|\xi_n-\xi|\geqslant\varepsilon\}$, $A^\varepsilon=\overline{\lim}A_n^\varepsilon\equiv\bigcap_{n=1}^\infty\bigcup_{k\geqslant n}A_k^\varepsilon$. 那么,

$$\{\xi_n\nrightarrow\xi\}=\bigcup_{\varepsilon>0}A^\varepsilon=\bigcup_{m=1}^\infty A^{1/m}.$$

因为

$$\mathbf{P}(A^\varepsilon)=\lim_n\mathbf{P}\left(\bigcup_{k\geqslant n}A_k^\varepsilon\right),$$

所以命题 a) 的结论由下面一系列等价关系得出:

$$\mathbf{P}\{\omega:\xi_n\nrightarrow\xi\}=0\Leftrightarrow\mathbf{P}\left(\bigcup_{\varepsilon>0}A^\varepsilon\right)=0\Leftrightarrow\mathbf{P}\left(\bigcup_{m=1}^\infty A^{1/m}\right)=0\Leftrightarrow$$

$$\Leftrightarrow\mathbf{P}(A^{1/m})=0,m\geqslant 1\Leftrightarrow\mathbf{P}(A^\varepsilon)=0,\varepsilon>0\Leftrightarrow$$

$$\Leftrightarrow\mathbf{P}\left(\bigcup_{k\geqslant n}A_k^\varepsilon\right)=0,n\to\infty,\varepsilon>0\Leftrightarrow\mathbf{P}\left\{\sup_{k\geqslant n}|\xi_k-\xi|\geqslant\varepsilon\right\}\to 0,n\to\infty,\varepsilon>0.$$

b) 记

$$B_{k,l}^\varepsilon=\{\omega:|\xi_k-\xi_l|\geqslant\varepsilon\},B^\varepsilon=\bigcap_{n=1}^\infty\bigcup_{\substack{k\geqslant n\\ l\geqslant n}}B_{k,l}^\varepsilon.$$

那么, $\{\omega:\{\xi_n(\omega)\}$ 不是基本的$\}=\bigcup_{\varepsilon>0}B^\varepsilon$, 像 a) 一样可以证明 $\mathbf{P}\{\omega:\{\xi_n(\omega)\}$ 不是基本的$\}=0$ 与 (6) 式等价. 而由下面明显的不等式, 可见 (6) 式与 (7) 式等价:

$$\sup_{k\geqslant 0}|\xi_{n+k}-\xi_n|\leqslant\sup_{k\geqslant n,l\geqslant n}|\xi_{n+k}-\xi_{n+l}|\leqslant 2\sup_{k\geqslant 0}|\xi_{n+k}-\xi_n|.$$ □

系 由于

$$\mathbf{P}\left\{\sup_{k\geqslant n}|\xi_k-\xi|\geqslant\varepsilon\right\}=\mathbf{P}\left(\bigcup_{k\geqslant n}\{|\xi_k-\xi|\geqslant\varepsilon\}\right)\leqslant\sum_{k\geqslant n}\mathbf{P}\{|\xi_k-\xi|\geqslant\varepsilon\},$$

可见 $\xi_n\to\xi(\mathbf{P}-\text{a.c.})$ 的充分条件是: 对于任意 $\varepsilon>0$, 有

$$\sum_{k=1}^\infty\mathbf{P}\{|\xi_k-\xi|\geqslant\varepsilon\}<\infty. \tag{8}$$

现在关于条件 (8) 应当指出, 推导该式的论断可以用来建立如下简单但十分重要的结果, 在研究依概率 1 成立的性质时是基本工具.

设 $A_1,A_2,\cdots$ 是 $\mathscr{F}$ 中的某一事件序列. 我们 (见 §1 表 2 – 1) 曾使用记号 {无限多个 A_n} 表示事件 $\overline{\lim}A_n$ 为 "在 $A_1,A_2,\cdots$ 中有无限多个出现".

引理 1 (博雷尔 – 坎泰利 [E. Borel-F. P. Cantelli]) a) 如果 $\sum \mathbf{P}(A_n) < \infty$, 则概率 $\mathbf{P}\{$无限多个$A_n\} = 0$.

b) 如果 $\sum \mathbf{P}(A_n) = \infty$ 且事件 $A_1, A_2, \cdots$ 独立, 则概率 $\mathbf{P}\{$无限多个 $A_n\} = 1$.

证明 a) 根据定义

$$\{\text{无限多个 } A_n\} = \overline{\lim} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k.$$

因此,

$$\mathbf{P}\{\text{无限多个 } A_n\} = \mathbf{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geqslant n} A_k\right) = \lim \mathbf{P}\left(\bigcup_{k \geqslant n} A_k\right) \leqslant \sum_{k \geqslant n} \mathbf{P}(A_k),$$

由此命题 a) 得证.

b) 如果 $A_1, A_2, \cdots$ 独立, 则 $\overline{A}_1, \overline{A}_2, \cdots$ 也独立. 那么, 对于任意 $N \geqslant n$, 有

$$\mathbf{P}\left(\bigcap_{k=n}^{N} \overline{A}_k\right) = \prod_{k=n}^{N} \mathbf{P}(\overline{A}_k),$$

由此易见

$$\mathbf{P}\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) = \prod_{k=n}^{\infty} \mathbf{P}(\overline{A}_k). \tag{9}$$

由不等式 $\ln(1-x) \leqslant -x, 0 \leqslant x < 1$, 可得,

$$\ln \prod_{k=n}^{\infty} [1 - \mathbf{P}(A_k)] = \sum_{k=n}^{\infty} \ln[1 - \mathbf{P}(A_k)] \leqslant -\sum_{k=n}^{\infty} \mathbf{P}(A_k) = -\infty.$$

从而, 对于任意 n,

$$\mathbf{P}\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) = 0,$$

即 $\mathbf{P}\{$无限多个 $A_n\} = 1$. □

系 1 如果 $A_n^{\varepsilon} = \{\omega : |\xi_n - \xi| \geqslant \varepsilon_n\}$, 则条件 (8) 表示

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n^{\varepsilon}) < \infty, \quad \varepsilon > 0,$$

而根据博雷尔 – 坎泰利引理 $\mathbf{P}(A^{\varepsilon}) = 0, \varepsilon > 0$, 其中 $A^{\varepsilon} = \overline{\lim} A_n^{\varepsilon} (= \{$无限多个$A_n^{\varepsilon}\})$. 于是, 有

$$\sum_{k=1}^{\infty} \mathbf{P}\{|\xi_k - \xi| \geqslant \varepsilon\} < \infty, \varepsilon > 0 \Rightarrow \mathbf{P}(A^{\varepsilon}) = 0, \varepsilon > 0 \Leftrightarrow \mathbf{P}\{\omega : \xi_n \nrightarrow \xi\} = 0$$

而关于这一点上面已经指出.

系 2 设 $\{\varepsilon_n\}_{n\geqslant 1}$ 是正数序列,$\varepsilon_n \downarrow 0, n\to\infty$. 那么, 如果

$$\sum_{n=1}^{\infty} \mathbf{P}\{|\xi_n-\xi|\geqslant \varepsilon_n\}<\infty, \tag{10}$$

则 $\xi_n \xrightarrow{\text{a.c.}} \xi$.

事实上, 设 $A_n=\{|\xi_n-\xi|\geqslant\varepsilon_n\}$. 那么, 根据博雷尔－坎泰利引理 $\mathbf{P}\{$无限多个 $A_n\}=0$. 对于几乎每一个结局 $\omega\in\Omega$, 存在 $N=N(\omega)$, 使对于 $n\geqslant N(\omega)$, 有 $|\xi_n(\omega)-\xi(\omega)|<\varepsilon_n$. 由于 $\varepsilon_n\downarrow 0$, 可见对于几乎一切 $\omega\in\Omega, \xi_n(\omega)\to\xi(\omega)$.

4. 各种收敛性的蕴涵关系

定理 2 有如下蕴涵关系:

$$\xi_n \xrightarrow{\text{a.c.}} \xi \Rightarrow \xi_n \xrightarrow{\mathbf{P}} \xi, \tag{11}$$

$$\xi_n \xrightarrow{L^p} \xi \Rightarrow \xi_n \xrightarrow{\mathbf{P}} \xi, \quad p>0, \tag{12}$$

$$\xi_n \xrightarrow{\mathbf{P}} \xi \Rightarrow \xi_n \xrightarrow{d} \xi. \tag{13}$$

证明 由 (5) 式可见命题 (11) 成立; 由切比雪夫不等式, 可见蕴涵关系 (12) 成立.

现在证明蕴涵关系 (13). 设 $f(x)$ 是连续函数, $|f(x)|\leqslant c, \varepsilon>0$, 而 N 满足 $\mathbf{P}\{|\xi|>N\}\leqslant\varepsilon/(4c)$.

选择 $\delta>0$, 使得对于一切 $|x|\leqslant N$ 和 $|x-y|\leqslant\delta$, 不等式

$$|f(x)-f(y)|\leqslant\frac{\varepsilon}{2}$$

成立.

那么, (对照第一章 §5 第 5 小节, 维尔斯特拉斯定理的"概率的"证明):

$$\begin{aligned}
\mathbf{E}|f(\xi_n)-f(\xi)| &= \mathbf{E}[|f(\xi_n)-f(\xi)|;|\xi_n-\xi|\leqslant\delta,|\xi|\leqslant N] \\
&\quad +\mathbf{E}[|f(\xi_n)-f(\xi)|;|\xi_n-\xi|\leqslant\delta,|\xi|>N] \\
&\quad +\mathbf{E}(|f(\xi_n)-f(\xi)|;|\xi_n-\xi|>\delta) \\
&\leqslant \varepsilon/2+\varepsilon/2+2c\mathbf{P}[|\xi_n-\xi|>\delta]=\varepsilon+2c\mathbf{P}[|\xi_n-\xi|>\delta].
\end{aligned}$$

由于 $\mathbf{P}(|\xi_n-\xi|>\delta)\to 0$, 可见对于充分大的 $n, \mathbf{E}|f(\xi_n)-f(\xi)|\leqslant 2\varepsilon$. 由于 $\varepsilon>0$ 是任意的, 可见蕴涵关系 (13) 得证. □

现在举一些例子, 其中有的说明 (11),(12) 式中相反的蕴涵关系一般不成立.

例 1 $(\xi_n \xrightarrow{\mathbf{P}} \xi \nRightarrow \xi_n \xrightarrow{\text{a.c.}} \xi,\ \xi_n \xrightarrow{L^p} \xi \nRightarrow \xi_n \xrightarrow{\text{a.c.}} \xi)$. 设 $\Omega=[0,1], \mathscr{F}=\mathscr{B}[0,1], \mathbf{P}$ 是 $[0,1]$ 上的勒贝格测度. 设

$$A_n^i=\left[\frac{i-1}{n},\frac{i}{n}\right], \xi_n^i=I_{A_n^i}(\omega), i=1,2,\cdots,n, n\geqslant 1.$$

那么, 随机变量序列

$$\{\xi_1^1; \xi_2^1, \xi_2^2; \xi_3^1, \xi_3^2, \xi_3^3; \cdots\}$$

依概率收敛, 也对 $p>0$ 阶平均收敛, 但是它在任何点 $\omega \in [0,1]$ 都不收敛.

例 2 $(\xi_n \xrightarrow{\mathbf{P}} \xi \Leftarrow \xi_n \xrightarrow{\text{a.c.}} \xi \nRightarrow \xi_n \xrightarrow{L^p} \xi (p>0))$. 仍设 $\Omega = [0,1], \mathscr{F} = \mathscr{B}[0,1], \mathbf{P}$ 是 $[0,1]$ 上的勒贝格测度, 而

$$\xi_n(\omega) = \begin{cases} \mathrm{e}^n, & 若\ 0 \leqslant \omega \leqslant 1/n, \\ 0, & 若\ \omega > 1/n. \end{cases}$$

那么, 序列 $\{\xi_n\}$ 依概率 1(故依概率) 收敛于 0, 但是对于任何 $p>0$,

$$\mathbf{E}|\xi_n|^p = \frac{\mathrm{e}^{np}}{n} \to \infty, \quad n \to \infty.$$

例 3 $(\xi_n \xrightarrow{L^p} \xi \nRightarrow \xi_n \xrightarrow{\text{a.c.}} \xi)$. 设 $\{\xi_n\}$ 是独立随机变量序列, 且

$$\mathbf{P}\{\xi_n = 1\} = p_n, \quad \mathbf{P}\{\xi_n = 0\} = 1 - p_n.$$

那么, 不难证明:

$$\xi_n \xrightarrow{\mathbf{P}} 0 \Leftrightarrow p_n \to 0, \quad n \to \infty, \tag{14}$$

$$\xi_n \xrightarrow{L^p} 0 \Leftrightarrow p_n \to 0, \quad n \to \infty, \tag{15}$$

$$\xi_n \xrightarrow{\text{a.c.}} 0 \Leftrightarrow \sum_{n=1}^{\infty} \xi_n < \infty. \tag{16}$$

特别, 当 $p_n = 1/n$ 时, 对于任何 $p>0, \xi_n \xrightarrow{L^p} 0$ 然而 ξ_n 并不几乎处处收敛于 0.

下面的定理涉及一个重要的情形: 由几乎处处收敛可以导出 L^1 平均收敛.

定理 3 *设 $\{\xi_n\}_{n\geqslant 1}$ 是非负随机变量序列, 且 $\xi_n \xrightarrow{\text{a.c.}} \xi, \mathbf{E}\xi_n \to \mathbf{E}\xi < \infty$. 那么,*

$$\mathbf{E}|\xi_n - \xi| \to 0, \quad n \to \infty. \tag{17}$$

证明 对于充分大的 n, 有 $\mathbf{E}\xi_n < \infty$. 因此

$$\begin{aligned} \mathbf{E}|\xi - \xi_n| &= \mathbf{E}(\xi - \xi_n) I_{\{\xi \geqslant \xi_n\}} + \mathbf{E}(\xi_n - \xi) I_{\{\xi_n > \xi\}} \\ &= 2\mathbf{E}(\xi - \xi_n) I_{\{\xi \geqslant \xi_n\}} + \mathbf{E}(\xi_n - \xi). \end{aligned}$$

由于 $0 \leqslant \mathbf{E}(\xi - \xi_n) I_{\{\xi \geqslant \xi_n\}} \leqslant \xi$, 则由控制收敛定理知 $\lim_n \mathbf{E}(\xi - \xi_n) I_{\{\xi \geqslant \xi_n\}} = 0$, 于是由假设 $\mathbf{E}\xi_n \to \mathbf{E}\xi$ 证得 (17) 式. □

注 如果将依概率 1 收敛换成依概率收敛, 则控制收敛定理 (§6 定理 3) 仍然成立 (见练习题 1). 因此, 在定理 3 中收敛 “$\xi_n \xrightarrow{\text{a.c.}} \xi$” 可以换成收敛 “$\xi_n \xrightarrow{\mathbf{P}} \xi$”.

5. 柯西收敛准则　由数学分析知, 任意基本数列 $\{x_n\}, x_n \in \mathbb{R}$, 都是收敛的 (柯西准则). 我们对于随机变量序列引进类似的结果.

定理 4 (柯西几乎必然收敛准则)　*随机变量序列 $\{\xi_n\}$ 依概率 1 收敛 (于某一随机变量 ξ) 的充分必要条件是 $\{\xi_n\}$ 依概率 1 为基本序列.*

证明　若 $\xi_n \xrightarrow{\text{a.c.}} \xi$, 则

$$\sup_{k\geqslant n, l\geqslant n} |\xi_k - \xi_l| \leqslant \sup_{k\geqslant n} |\xi_k - \xi| + \sup_{l\geqslant n} |\xi_l - \xi|,$$

由此 (见定理 1) 可得定理的必要性.

现在假设 $\{\xi_n\}$ 依概率 1 为基本序列. 记 $\mathscr{N} = \{\omega : \{\xi_n(\omega)\}$ *不是基本序列*$\}$. 那么, 对于 $\omega \in \Omega\backslash\mathscr{N}$, 数列 $\{\xi_n(\omega)\}$ 是基本的, 故根据数列的柯西定理, 数列 $\{\xi_n(\omega)\}$ 有极限 $\lim \xi_n(\omega)$. 设

$$\xi(\omega) = \begin{cases} \lim \xi_n(\omega), & \text{若 } \omega \in \Omega\backslash\mathscr{N}, \\ 0, & \text{若 } \omega \in \mathscr{N}. \end{cases} \tag{18}$$

于是, 所定义的函数 $\xi(\omega)$ 是随机变量, 因而显然 $\xi_n \xrightarrow{\text{a.c.}} \xi$. □

定理 5　*设 $\{\xi_n\}$ 对于依概率收敛是基本随机变量序列, 则从中可以分离出依概率 1 收敛的子序列 $\{\xi_{n_k}\}$.*

证明　设 $\{\xi_n\}$ 对于依概率收敛是基本随机变量序列. 由于定理 4, 只需证明从中可以分出几乎处处收敛的子序列.

设 $n_1 = 1$, 由归纳法确定一个 n_k, 使之对于任意 $s \geqslant n, t \geqslant n$, 是满足

$$\mathbf{P}\{|\xi_t - \xi_s| > 2^{-k}\} < 2^{-k}$$

的最小 $n > n_{k-1}$. 那么,

$$\sum_k \mathbf{P}\{|\xi_{n_{k+1}} - \xi_{n_k}| > 2^{-k}\} < \sum_k 2^{-k} < \infty,$$

而由博雷尔 – 坎泰利引理

$$\mathbf{P}\{|\xi_{n_{k+1}} - \xi_{n_k}| > 2^{-k}, \text{无限多个 } k\} = 0.$$

因此, 依概率 1, 有

$$\sum_k |\xi_{n_{k+1}} - \xi_{n_k}| < \infty.$$

设 $\mathscr{N} = \{\omega : \sum |\xi_{n_{k+1}} - \xi_{n_k}| = \infty\}$. 那么, 如果设

$$\xi(\omega) = \begin{cases} \xi_{n_1}(\omega) + \sum_{k=1}^{\infty} [\xi_{n_{k+1}}(\omega) - \xi_{n_k}(\omega)], & \text{若 } \omega \in \Omega\backslash\mathscr{N}, \\ 0, & \text{若 } \omega \in \mathscr{N}, \end{cases}$$

则 $\xi_{n_k} \xrightarrow{\text{a.c.}} \xi$.

如果原序列依概率收敛, 则它按依概率为基本的 (见下面的 (19) 式), 从而这正是所讨论的情形. □

定理 6 (柯西依概率收敛准则) 随机变量序列 $\{\xi_n\}$ 依概率收敛的充分必要条件是, $\{\xi_n\}$ 按依概率收敛是基本随机变量序列.

证明 如果 $\xi_n \xrightarrow{\mathbf{P}} \xi$, 则

$$\mathbf{P}\{|\xi_n - \xi_m| \geqslant \varepsilon\} \leqslant \mathbf{P}\{|\xi_n - \xi| \geqslant \varepsilon/2\} + \mathbf{P}\{|\xi_m - \xi| \geqslant \varepsilon/2\}, \tag{19}$$

即 $\{\xi_n\}$ 按依概率收敛是基本随机变量序列.

相反, 若 $\{\xi_n\}$ 按依概率收敛是基本随机变量序列, 则根据定理 5, 存在随机变量子序列 $\{\xi_{n_k}\}$ 和随机变量 ξ, 使 $\xi_{n_k} \xrightarrow{\text{a.c.}} \xi$. 那么,

$$\mathbf{P}\{|\xi_n - \xi| \geqslant \varepsilon\} \leqslant \mathbf{P}\{|\xi_n - \xi_{n_k}| \geqslant \varepsilon/2\} + \mathbf{P}\{|\xi_{n_k} - \xi| \geqslant \varepsilon/2\},$$

即 $\xi_n \xrightarrow{\mathbf{P}} \xi$. □

关于 $p(>0)$ 阶平均收敛, 我们首先对空间 L^p 作若干说明.

以 $L^p = L^p(\Omega, \mathscr{F}, \mathbf{P})$ 表示满足

$$\mathbf{E}|\xi|^p \equiv \int_\Omega |\xi|^p \ \mathrm{d}\mathbf{P} < \infty$$

的随机变量 $\xi = \xi(\omega)$ 的空间. 假设 $p \geqslant 1$, 并设

$$\|\xi\|_p = (\mathbf{E}|\xi|^p)^{1/p}.$$

显然

$$\|\xi\|_p \geqslant 0, \tag{20}$$

$$\|c\xi\|_p = |c|\|\xi\|_p, \quad c \text{ 是常数}, \tag{21}$$

由闵可夫斯基不等式 (§6 (31) 式), 有

$$\|\xi + \eta\|_p \leqslant \|\xi\|_p + \|\eta\|_p. \tag{22}$$

这样, 根据泛函分析中熟知的定义, 定义在 L^p 上并满足条件 (20)~(22) 式的函数 $\|\cdot\|_p$ (对于 $p > 1$) 称做*半范数*.

为使之成为*范数*尚需具有如下性质

$$\|\xi\|_p = 0 \Rightarrow \xi = 0, \tag{23}$$

当然这一般并不成立, 因为根据性质 $\mathbf{H}$(§6), 只是可以断定 ξ 几乎必然为 0, 而不是恒等于 0.

鉴于这种情况, 需要用略有不同的观点对待空间 L^p. 具体地说, 将每一个随机变量 $\xi \in L^p$ 和 L^p 中与之等价的随机变量类 $[\xi]$ 相联系 (称 ξ 和 η 等价, 如果 $\mathbf{P}\{\xi = \eta\} = 1$). 不难证明, 等价性具有自反性、对称性和传递性, 说明线性空间 L^p

可以分割为两两不相交的等价随机变量类. 如果以 $[L^p]$ 表示 L^p 中等价的随机变量类 $[\xi]$ 的全体, 并且定义:

$$[\xi]+[\eta]=[\xi+\eta],$$
$$a[\xi]=[a\xi], a \text{ 是常数},$$
$$\|[\xi]\|_p=\|\xi\|_p,$$

则 $[L^p]$ 就是线性赋范空间.

泛函分析中关于空间 $[L^p]$ 的元素, 通常称做"函数", 而不是"等价函数类". 因此, 我们以后不再使用记号 $[L^p]$, 而正是把 L^p 理解为等价函数类的集合, 像以前一样简单地称元素, 函数, 随机变量 …… .

泛函分析重要的结果之一, 就是要证明空间 $L^p, p\geqslant 1$, 是*完备的*, 即任何基本序列都是收敛的. 下面用概率的语言表述并证明这一结果.

定理 7 (柯西 $p(p\geqslant 1)$ 阶平均收敛准则) *空间 L^p 中的随机变量序列 $\{\xi_n\}$, $p(p\geqslant 1)$ 阶平均收敛于取值于 L^p 的随机变量的充分必要条件是, $\{\xi_n\}$ 是 p 阶平均收敛的基本随机变量序列.*

证明 由闵可夫斯基不等式, 可见必要性成立. 设 $\{\xi_n\}$ 是基本随机变量序列 $(\|\xi_n-\xi_m\|_p\to 0, n,m\to\infty)$. 仿照定理 5 的证明, 选择一子序列 $\{\xi_{n_k}\}$ 使 $\xi_{n_k}\xrightarrow{\text{a.c.}}\xi$, 其中 ξ 是某一 $\|\xi\|_p<\infty$ 的随机变量.

设 $n_1=1$, 由归纳法确定一个 n_k, 使之对于任意 $s\geqslant n, t\geqslant n$, 是满足

$$\|\xi_t-\xi_s\|_p<2^{-2k}$$

的最小 $n>n_{k-1}$. 记

$$A_k=\{\omega:|\xi_{n_{k+1}}-\xi_{n_k}|\geqslant 2^{-k}\}.$$

那么, 由切比雪夫不等式, 有

$$\mathbf{P}(A_k)\geqslant\frac{\mathbf{E}|\xi_{n_{k+1}}-\xi_{n_k}|^p}{2^{-kp}}\leqslant\frac{2^{-2kp}}{2^{-kp}}=2^{-kp}\leqslant 2^{-k}.$$

如同定理 5, 由此可见存在随机变量 ξ, 使 $\xi_{n_k}\xrightarrow{\text{a.c.}}\xi$.

由此可见当 $n\to\infty$ 时 $\|\xi_n-\xi\|_p\to 0$. 为此固定 $\varepsilon>0$ 并选择 $N=N(\varepsilon)$, 使对于 $n\geqslant N, m\geqslant N$, 有 $\|\xi_n-\xi_m\|_p^p<\varepsilon$. 那么, 对于任意固定的 $n\geqslant N$, 由法图引理 (§6), 有

$$\begin{aligned}\mathbf{E}|\xi_n-\xi|^p&=\mathbf{E}\left\{\lim_{n_k\to\infty}|\xi_n-\xi_{n_k}|^p\right\}=\mathbf{E}\left\{\varliminf_{n_k\to\infty}|\xi_n-\xi_{n_k}|^p\right\}\\&\leqslant\varliminf_{n_k\to\infty}\mathbf{E}|\xi_n-\xi_{n_k}|^p=\varliminf_{n_k\to\infty}|\xi_n-\xi_{n_k}|_p^p\leqslant\varepsilon.\end{aligned}$$

从而, 当 $n\to\infty$ 时 $|\xi_n-\xi|^p\to 0$. 由于 $\xi=(\xi-\xi_n)+\xi_n$, 故由闵可夫斯基不等式可见 $\mathbf{E}|\xi|^p<\infty$. □

注 1 按照泛函分析的术语, 完备赋范线性空间称为**巴拿赫** (S. Banach) 空间. 这样, 空间 $L^p, p \geqslant 1$, 是巴拿赫空间.

注 2 如果 $0 < p < 1$, 则 $\|\xi\|_p = (\mathbf{E}|\xi|^p)^{1/p}$ 不满足三角形不等式 (22), 从而不是范数. 然而, 空间 (等价类) $L^p, 0 < p < 1$, 关于度量 $d(\xi, \eta) \equiv \mathbf{E}|\xi - \eta|^p$ **是完备的**.

注 3 设随机变量 $\xi = \xi(\omega)$ 满足条件: $\|\xi\|_\infty < \infty$, 其中 $\|\xi\|_\infty$ 决定于

$$\|\xi\|_\infty \equiv \text{ess sup}\,|\xi| \equiv \inf\{0 \leqslant c \leqslant \infty : \mathbf{P}(|\xi| > c) = 0\},$$

称做 ξ 的**本质上确界**. 记 $L^\infty = L^\infty(\Omega, \mathscr{F}, \mathbf{P})$ 是随机变量 $\xi = \xi(\omega)$ 的 (等价类) 空间.

函数 $\|\cdot\|_\infty$ 是范数, 且空间 L^∞ 关于此范数是完备的.

6. 练习题

1. 利用定理 5, 证明 §6 定理 3 和定理 4 的命题: “依概率 1 收敛” 可以换成 “依概率收敛”.

2. 证明 L^∞ 是完备空间.

3. 证明, 若 $\xi_n \xrightarrow{\mathbf{P}} \xi$ 且 $\eta_n \xrightarrow{\mathbf{P}} \eta$, 则 ξ 和 η 等价 (即 $\mathbf{P}\{\xi \neq \eta\} = 0$).

4. 设 $\xi_n \xrightarrow{\mathbf{P}} \xi$ 和 $\eta_n \xrightarrow{\mathbf{P}} \eta$, 且 ξ 和 η 等价. 证明对于任意 $\varepsilon > 0$, 有

$$\mathbf{P}\{|\xi_n - \eta_n| \geqslant \varepsilon\} \to 0, n \to \infty$$

5. 设 $\xi_n \xrightarrow{\mathbf{P}} \xi$ 和 $\eta_n \xrightarrow{\mathbf{P}} \eta$. 证明, 如果 $\varphi = \varphi(x, y)$ 是连续函数, 则 $\varphi(\xi_n, \eta_n) \xrightarrow{\mathbf{P}} \varphi(\xi, \eta)$ (**斯鲁斯基** [E. E. Слуцкий] **引理**).

6. 设 $(\xi_n - \xi)^2 \xrightarrow{\mathbf{P}} 0$. 证明 $\xi_n^2 \xrightarrow{\mathbf{P}} \xi^2$.

7. 证明, 若 $\xi_n \xrightarrow{d} C$, 其中 C 为常数, 则有依概率收敛:

$$\xi_n \xrightarrow{d} C \quad \Rightarrow \quad \xi_n \xrightarrow{\mathbf{P}} C.$$

8. 设序列 $\{\xi_n\}_{n\geqslant 1}$ 满足条件: 对于某个 $p > 0, \sum_{n=1}^{\infty} \mathbf{E}|\xi_n|^p < \infty$, 证明 $\xi_n \xrightarrow{\mathbf{P}-\text{a.c.}} 0$.

9. 设 $\{\xi_n\}_{n\geqslant 1}$ 是同分布随机变量序列, 证明

$$\mathbf{E}|\xi_1| < \infty \Leftrightarrow \sum_{n=1}^{\infty} \mathbf{P}\{|\xi_1| > \varepsilon n\} < \infty, \varepsilon > 0$$

$$\Leftrightarrow \sum_{n=1}^{\infty} \mathbf{P}\left\{\left|\frac{\xi_n}{n}\right| > \varepsilon\right\} < \infty, \varepsilon > 0 \Rightarrow \frac{\xi_n}{n} \xrightarrow{\mathbf{P}-\text{a.c.}} 0.$$

10. 设 $\{\xi_n\}_{n\geqslant 1}$ 是一随机变量序列. 假设存在随机变量 ξ 和子序列 $\{n_k\}$ 使 $\xi_{n_k} \xrightarrow{\mathbf{P}-\text{a.c.}} \xi$ 和 $\max\limits_{n_{k-1} < l \leqslant n_k} |\xi_l - \xi_{n_{k-1}}| \xrightarrow{\mathbf{P}-\text{a.c.}} \xi, k \to \infty$. 证明 $\xi_n \to \xi(\mathbf{P}-\text{a.c.})$.

11. 在随机变量的集合中引进 “度量 d” 如下:

$$d(\xi, \eta) = \mathbf{E}\frac{|\xi - \eta|}{1 + |\xi - \eta|},$$

并且认为几乎处处相等的随机变量相同. 证明 $d = d(\xi, \eta)$ 确实是度量, 并且依概率收敛等价于在此度量下的收敛.

12. 证明在随机变量的集合中, 不存在与几乎处处收敛等价的度量.

13. 设 $X_1 \leqslant X_2 \leqslant \cdots$, 且 $X_n \xrightarrow{\mathbf{P}} X$, 证明 $X_n \xrightarrow{\mathbf{P}-\text{a.c.}} X$.

14. 设 $X_n \xrightarrow{\mathbf{P}-\text{a.c.}} X$, 则 $\overline{X}_n \xrightarrow{\mathbf{P}-\text{a.c.}} X$ (切萨罗 [E. Cesaro] 求和法) 其中

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

举例说明 $\mathbf{P}-$ 几乎处处收敛不能换成依概率收敛.

15. 设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, $X_n \xrightarrow{\mathbf{P}} X$ 证明, 如果 $\mathbf{P}$ 是原子测度, 则也依概率 1 有 $X_n \to X$. (若对于任意 $B \in \mathscr{F}$, 要么 $\mathbf{P}(B \bigcap A) = \mathbf{P}(A)$, 要么 $\mathbf{P}(B \bigcap A) = 0$, 则集合 $A \in \mathscr{F}$ 称做 $\mathbf{P}-$ *原子*. 测度 $\mathbf{P}$ 称做 $\mathbf{P}-$ *原子的*, 如果存在可数个不相交的 $\mathbf{P}-$ 原子族 $\{A_n\}$, 使 $\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = 1$.)

16. 根据空间 (第一) 博雷尔 – 坎泰利引理, 如果对于任意 $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbf{P}\{|\xi_n| > \varepsilon\} < \infty,$$

则序列 $\xi_n \xrightarrow{\mathbf{P}-\text{a.c.}} 0$. 举例说明, 在 $\sum_{n=1}^{\infty} \mathbf{P}(|\xi_n| > \varepsilon) = \infty (\varepsilon > 0)$ 的条件下, $\xi_n \xrightarrow{\mathbf{P}-\text{a.c.}} 0$ 也可能成立.

17. (*第二博雷尔 – 坎泰利引理*). 设 $\Omega = (0,1), \mathscr{B} = \mathscr{B}((0,1)), \mathbf{P}$ 是勒贝格测度. 考虑事件 $A_n = (0, 1/n)$. 证明 $\sum \mathbf{P}(A_n) = \infty$, 但是 $(0,1)$ 上的每一个 ω 最多只可能属于有限个集合 $A_1, \cdots, A_{[1/\omega]}$, 即 $\mathbf{P}\{\text{无限多个} A_n\} = 0$.

18. 举一随机变量序列的例子 $\{\xi_n\}$, 使之依概率 1 有 $\limsup \xi_n = \infty, \liminf \xi_n = -\infty$, 然而存在随机变量 $\eta: \xi_n \xrightarrow{\mathbf{P}} \eta$.

19. 设 Ω 是有限或可数集合, 证明, 若 $\xi_n \xrightarrow{\mathbf{P}} \xi$, 则 $\xi_n \xrightarrow{\mathbf{P}-\text{a.c.}} \xi$.

20. 设 $A_1, A_2, \cdots$ 是独立事件序列且 $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$. 证明, 对于 $S_n = \sum_{k=1}^{n} I(A_k)$, 满足 “第二博雷尔 – 坎泰利引理” 的条件:

$$\lim_n \frac{S_n}{\mathbf{E}S_n} = 1 \quad (\mathbf{P}-\text{a.c.}).$$

21. 设 $(X_n)_{n\geqslant 1}$ 和 $(Y_n)_{n\geqslant 1}$ 是两个随机变量序列, 且其一切有限维分布相同 $(F_{X_1,\cdots,X_n} = F_{Y_1,\cdots,Y_n}, n \geqslant 1)$. 证明, 若 $X_n \xrightarrow{\mathbf{P}} X$, 则 $Y_n \xrightarrow{\mathbf{P}} Y$, 其中 Y 是与 X 同分布的随机变量.

22. 设 $(X_n)_{n\geqslant 1}$ 是独立随机变量序列, 且 $X_n \xrightarrow{\mathbf{P}} X$, 其中 X 是某一随机变量. 证明 X 是退化随机变量.

23. 证明, 对于每一个随机变量序列 $\xi_1, \xi_2, \cdots$, 存在这样的数列 $a_1, a_2, \cdots$, 使 $\xi_n/a_n \xrightarrow{\mathbf{P}-\text{a.c.}} 0$.

24. 设 $\xi_1, \xi_2, \cdots$ 是随机变量序列, $S_n = \xi_1 + \cdots + \xi_n, n \geqslant 1$. 证明 $\{S_n \to\}$ —— 使级数 $\sum_{k\geqslant 1} \xi_k(\omega)$ 收敛的 $\omega \in \Omega$ 的集合, 可以表示为如下形式:

$$\{S_n \to\} = \bigcap_{N\geqslant 1} \bigcup_{m\geqslant 1} \bigcap_{k\geqslant m} \left\{\sup_{l\geqslant k} |S_l - S_k| \leqslant N^{-1}\right\}.$$

相应地, 若级数 $\sum_{k\geqslant 1} \xi_k(\omega)$ 发散, 则

$$\{S_n \nrightarrow\} = \bigcup_{N\geqslant 1} \bigcap_{m\geqslant 1} \bigcup_{k\geqslant m} \left\{\sup_{l\geqslant k} |S_l - S_k| > N^{-1}\right\}.$$

25. 证明第二博雷尔 – 坎泰利引理的如下类型 (引理 1 的命题 b)): 假设事件 $A_1, A_2, \cdots$ (未必独立) 满足条件

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty \quad 和 \quad \liminf_n \frac{\sum\limits_{i,k\leqslant n} \mathbf{P}(A_i \cap A_k)}{\left[\sum\limits_{1<k\leqslant n} \mathbf{P}(A_k)\right]^2} \leqslant 1,$$

则 $\mathbf{P}\{无限多个\ A_n\} = 1$.

26. 证明在第二博雷尔 – 坎泰利引理中, 代替 $A_1, A_2, \cdots$ 独立, 只需要求它们两两独立.

27. 证明 $0-1$ 律的如下形式 (对照第四章 §1 的 $0-1$ 律): 如果 $A_1, A_2, \cdots$ 两两独立, 则

$$\mathbf{P}\{无限多个\ A_n\} = \begin{cases} 0, & 若\ \sum \mathbf{P}(A_n) < \infty, \\ 1, & 若\ \sum \mathbf{P}(A_n) = \infty. \end{cases}$$

28. 设对于任意事件序列 $A_1, A_2, \cdots$, 满足 $\lim\limits_n \mathbf{P}(A_n) = 0$ 和 $\sum_n \mathbf{P}(A_n \bigcap \overline{A}_{n+1}) < \infty$, 证明 $\mathbf{P}\{无限多个\ A_n\} = 0$.

29. 证明, 若 $\sum_n \mathbf{P}\{|\xi_n| > n\} < \infty$, 则 $\limsup_n (|\xi_n|/n) \leqslant 1 (\mathbf{P}-\text{a.c.})$.

30. 设 $\xi_n \downarrow \xi (\mathbf{P}-\text{a.c.}), \mathbf{E}|\xi_n| < \infty, n \geqslant 1$ 且 $\inf_n \mathbf{E}\xi_n > -\infty$, 证明 $\xi_n \xrightarrow{L^1} \xi$, 即 $\mathbf{E}|\xi_n - \xi| \to 0$.

31. 利用博雷尔 – 坎泰利引理证明, 如果 $\mathbf{P}\{无限多个\ A_n\} = 1$, 当且仅当对于每一个集合 $A, \mathbf{P}(A) > 0, \sum_n \mathbf{P}(A \bigcap A_n) = \infty$.

32. 设事件 $A_1, A_2, \cdots$ 独立, 且对于一切 $n \geqslant 1, \mathbf{P}(A_n) < 1$, 证明 $\mathbf{P}\{无限多个\ A_n\} = 1$, 当且仅当 $\mathbf{P}(\bigcup A_n) = 1$.

33. 设 $X_1, X_2, \cdots$ 是独立随机变量, 且 $\mathbf{P}\{X_n = 0\} = 1/n, \mathbf{P}\{X_n = 1\} = 1 - 1/n$; 记 $E_n = \{X_n = 0\}$. 证明

$$\sum_{n=1}^{\infty} \mathbf{P}(E_n) = \infty, \quad \sum_{n=1}^{\infty} \mathbf{P}(\overline{E}_n) = \infty.$$

并由此得出结论: $\lim_n X_n(\mathbf{P} - \text{a.c.})$ 不存在.

34. 对于随机变量序列 $X_1, X_2, \cdots$, 证明 $X_n \xrightarrow{\mathbf{P}} 0$, 当且仅当对于某个 $r > 0$, 有

$$\mathbf{E}\frac{|X_n|^r}{1 + |X_n|^r} \to 0.$$

特别, 如果 $S_n = X_1 + \cdots + X_n$, 则

$$\frac{S_n - \mathbf{E}S_n}{n} \xrightarrow{\mathbf{P}} 0 \Leftrightarrow \mathbf{E}\frac{(S_n - \mathbf{E}S_n)^2}{n^2 + (S_n - \mathbf{E}S_n)^2} \to 0.$$

证明, 对于任意随机变量序列 $X_1, X_2, \cdots$, 有

$$\max_{1 \leqslant k \leqslant n} |X_k| \xrightarrow{\mathbf{P}} 0 \Rightarrow \frac{S_n}{n} \xrightarrow{\mathbf{P}} 0.$$

35. 假设对于独立同分布伯努利随机变量序列 $X_1, X_2, \cdots : \mathbf{P}\{X_k = \pm 1\} = 1/2$;

$$U_n = \sum_{k=1}^{n} \frac{X_k}{2^k}, \quad n \geqslant 1.$$

证明, $U_n \xrightarrow{\mathbf{P}-\text{a.c.}} U$, 其中 U 是在 $(-1, +1)$ 上均匀分布的随机变量.

§11. 具有有限二阶矩的随机变量的希尔伯特空间

1. 随机变量的希尔伯特 (D. Hilbert) 空间 在上面讨论的巴拿赫空间 $L^p, p \geqslant 1$ 中, 起重要作用的空间 $L^2 = L^2(\Omega, \mathscr{F}, \mathbf{P})$ 是具有有限二阶矩的等价随机变量类.

对于 $\xi, \eta \in L^2$, 记

$$(\xi, \eta) \equiv \mathbf{E}\xi\eta. \tag{1}$$

显然, 对于 $\xi, \eta, \zeta \in L^2$, 有

$$(a\xi + b\eta, \zeta) = a(\xi, \zeta) + b(\eta, \zeta), (\xi, \xi) \geqslant 0,$$

且

$$(\xi, \xi) = 0 \Rightarrow \xi = 0.$$

这样, (ξ, η) 是数量积. 关于由此数量积诱导的范数

$$\|\xi\| = (\xi, \xi)^{1/2}, \tag{2}$$

空间 L^2 是完备的 (证明见 §10). 因此, 按照泛函分析的术语, 引进数量积 (1) 的空间是 (具有有限二阶矩的) 随机变量的*希尔伯特空间*.

概率论中, 在研究仅由随机变量的前两阶矩决定的性质时, 广泛使用希尔伯特空间方法 ("L^2– 理论"). 这里, 我们仅限于介绍 (第六章) 叙述 L^2– 理论所必须的基本概念和事实.

2. 正交随机变量系 空间 L^2 中的两个随机变量 ξ 和 η 称做正交的 $(\xi \perp \eta)$, 如果其数量积 $(\xi,\eta) \equiv \mathbf{E}\xi\eta = 0$. 根据 §8, 随机变量 ξ 和 η 称做不相关的, 如果 $\operatorname{cov}(\xi,\eta) = 0$, 即如果

$$\mathbf{E}\xi\eta = \mathbf{E}\xi \times \mathbf{E}\eta.$$

由此可见, 对于均值为 0 的随机变量, *正交性与不相关性相同*.

如果对于任何随机变量 $\xi,\eta \in M(\xi \neq \eta), \xi \perp \eta$, 则系 $M \subseteq L^2$ 称做*正交随机变量系*.

此外, 如果对于一切 $\xi \in M$, 其范数 $\|\xi\| = 1$, 则 M 称做随机变量的*规范正交系*.

3. 最优线性估计量 设 $M = \{\eta_1, \cdots, \eta_n\}$ 是规范正交系, 而 ξ 是 L^2 中的某一随机变量. 在形如 $\sum_{i=1}^n a_i\eta_i$ 的线性估计类中, 求随机变量 ξ (在均方意义上) 的最优估计量 (对照 §8 的第 2 小节).

通过简单的计算, 可得

$$\begin{aligned}\mathbf{E}\left|\xi - \sum_{i=1}^n a_i\eta_i\right|^2 &\equiv \left\|\xi - \sum_{i=1}^n a_i\eta_i\right\|^2 = \left(\xi - \sum_{i=1}^n a_i\eta_i, \xi - \sum_{i=1}^n a_i\eta_i\right) \\ &= \|\xi\|^2 - 2\sum_{i=1}^n a_i(\xi,\eta_i) + \left(\sum_{i=1}^n a_i\eta_i, \sum_{i=1}^n a_i\eta_i\right) \\ &= \|\xi\|^2 - 2\sum_{i=1}^n a_i(\xi,\eta_i) + \sum_{i=1}^n a_i^2 \\ &= \|\xi\|^2 - \sum_{i=1}^n |(\xi,\eta_i)|^2 + \sum_{i=1}^n |a_i - (\xi,\eta_i)|^2 \\ &\geqslant \|\xi\|^2 - \sum_{i=1}^n |(\xi,\eta_i)|^2, \end{aligned} \tag{3}$$

其中用到了等式

$$a_i^2 - 2a_i(\xi,\eta_i) = |a_i - (\xi,\eta_i)|^2 - |(\xi,\eta_i)|^2.$$

由此可见, 对于一切实数 $a_1, \cdots, a_n$, 当 $a_i = (\xi,\eta_i), i = 1, 2, \cdots, n$ 时,

$$\mathbf{E}\left|\xi - \sum_{i=1}^n a_i\eta_i\right|^2$$

达到下确界.

这样, 由 $\eta_1, \cdots, \eta_n$ 对 ξ (在均方意义上) 的最优估计量为

$$\widehat{\xi} = \sum_{i=1}^{n} (\xi, \eta_i)\eta_i. \tag{4}$$

这时

$$\Delta \equiv \inf \mathbf{E} \left| \xi - \sum_{i=1}^{n} a_i \eta_i \right|^2 = \mathbf{E}|\xi - \widehat{\xi}|^2 = \|\xi\|^2 - \sum_{i=1}^{n} |(\xi, \eta_i)|^2 \tag{5}$$

(对照第一章 §4 (17) 式和 §8 (13) 式).

由 (3) 式亦可得如下贝塞尔 (F. W. Besser) 不等式: 如果 $M = \{\eta_1, \eta_2, \cdots\}$ 是规范正交系, 且 $\xi \in L^2$, 则

$$\sum_{i=1}^{\infty} |(\xi, \eta_i)|^2 \leqslant \|\xi\|^2, \tag{6}$$

其中达到等式当且仅当

$$\xi = \underset{n}{\text{l.i.m.}} \sum_{i=1}^{n} (\xi, \eta_i)\eta_i. \tag{7}$$

最优线性估计量 $\widehat{\xi}$ 常记作 $\widehat{\mathbf{E}}(\xi|\eta_1, \cdots, \eta_n)$, 并称为 ($\xi$ 关于 $\eta_1, \cdots, \eta_n$) 的**广义条件数学期望**.

关于这一术语有如下解释. 假如考虑随机变量 ξ 关于 $\eta_1, \cdots, \eta_n$ 的一切可能的估计量 $\varphi = \varphi(\eta_1, \cdots, \eta_n)$($\varphi$ 是博雷尔函数), 则 $\varphi^* = \mathbf{E}(\xi|\eta_1, \cdots, \eta_n)$, 即 ξ 关于 $\eta_1, \cdots, \eta_n$ 的条件数学期望是最优估计量 (对照 §8 的定理 1). 因此, 最优线性估计量类似地记作 $\widehat{\mathbf{E}}(\xi|\eta_1, \cdots, \eta_n)$, 并称为广义条件数学期望. 为此我们指出, 如果随机变量 $\eta_1, \cdots, \eta_n$ 构成高斯系 (见下面 §13), 则 $\mathbf{E}(\xi|\eta_1, \cdots, \eta_n)$ 与 $\widehat{\mathbf{E}}(\xi|\eta_1, \cdots, \eta_n)$ 相等.

下面说明估计量 $\widehat{\xi} = \widehat{\mathbf{E}}(\xi|\eta_1, \cdots, \eta_n)$ 的几何意义.

以 $\mathscr{L} = \mathscr{L}(\eta_1, \cdots, \eta_n)$ 表示规范正交随机变量 $\eta_1, \cdots, \eta_n$ 系生成的**线性流形** (即形如 $\sum_{i=1}^{n} a_i\eta_i (a_i \in \mathbb{R})$ 的随机变量系).

那么, 由以上的叙述可见, ξ 有如下 "正交分解":

$$\xi = \widehat{\xi} + (\xi - \widehat{\xi}), \tag{8}$$

其中 $\widehat{\xi} \in \mathscr{L}$, 而且在对于任意 $\lambda \in \mathscr{L}, \xi - \widehat{\xi} \perp \lambda$ 意义上 $\xi - \widehat{\xi} \perp \mathscr{L}$. 因此自然称 $\widehat{\xi}$ 为 ξ 在 $\mathscr{L}$ 上的**投影** (即 $\mathscr{L}$ 中 "最接近"ξ 的元素), $\xi - \widehat{\xi}$ 称做 $\mathscr{L}$ 的**垂线**.

4. 线性无关性 由随机变量 $\eta_1, \cdots, \eta_n$ 的规范正交性假设, 可以简单地找到由 $\eta_1, \cdots, \eta_n$ 对 ξ 的最优估计 $\widehat{\xi}$ (投影). 假如不要求规范正交性假设, 则情况比较复杂. 不过, 下面将要证明, 任意随机变量 $\eta_1, \cdots, \eta_n$ 的情形, 在一定意义上可以归为已经研究的规范正交变量的情形. 为简便计, 在以后的叙述中, 将假设所考虑的一切随机变量的均值为 0.

称随机变量 $\eta_1, \cdots, \eta_n$ 为线性无关的, 如果等式

$$\sum_{i=1}^{n} a_i \eta_i = 0 \quad (\mathbf{P} - \text{a.c.})$$

只有当所有 a_i 都等于零时才成立.

考虑协方差矩阵

$$\boldsymbol{R} \equiv \mathbf{E}\eta\eta^{\mathrm{T}},$$

其中 $\eta = (\eta_1, \cdots, \eta_n)$ 视为列向量. $\boldsymbol{R} \equiv \mathbf{E}\eta\eta^{\mathrm{T}}$ 是非负定对称矩阵, 如同在 §8 已经指出的, 存在正交矩阵 $\boldsymbol{Z}$, 可将 $\boldsymbol{R}$ 化为对角矩阵:

$$\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{Z} = \boldsymbol{D},$$

其中

$$\boldsymbol{D} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

的对角线上的非负元素 $d_i \geqslant 0 (i = 1, \cdots, n)$, 是矩阵 $\boldsymbol{R}$ 的特征数, 即特征方程 $\det(\boldsymbol{R} - \lambda\boldsymbol{E}) = 0$ 的根 λ ($\boldsymbol{E}$ 是单位矩阵).

如果随机变量 $\eta_1, \cdots, \eta_n$ 线性无关, 则克拉默 (J. P. Gram) 行列式 (即 $\det \boldsymbol{R}$) 不等于 0, 即全部 $d_i > 0 (i = 1, \cdots, n)$. 设

$$\boldsymbol{B} = \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_n} \end{pmatrix}$$

和

$$\beta = \boldsymbol{B}^{-1}\boldsymbol{Z}^{\mathrm{T}}\eta. \tag{9}$$

那么, 向量 β 的协方差矩阵

$$\mathbf{E}\beta\beta^{\mathrm{T}} = \boldsymbol{B}^{-1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{E}\eta\eta^{\mathrm{T}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{B}^{-1} = \boldsymbol{B}^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{Z}\boldsymbol{B}^{-1} = \boldsymbol{E},$$

从而向量 $\beta = (\beta_1, \cdots, \beta_n)$ 的分量是两两不相关的随机变量. 同样明显

$$\eta = (\boldsymbol{Z}\boldsymbol{B})\beta. \tag{10}$$

于是, 如果随机变量 $\eta_1, \cdots, \eta_n$ 线性无关, 则存在这样一个规范正交随机变量系 $\beta_1, \cdots, \beta_n$, 使 (9) 和 (10) 式成立. 这时

$$\mathscr{L}(\eta_1, \cdots, \eta_n) = \mathscr{L}(\beta_1, \cdots, \beta_n).$$

所介绍的得到规范正交随机变量系 $\beta_1,\cdots,\beta_n$ 的方法, 在一系列问题中显得不是很有用. 问题在于, 假如把 η_i 视为随机变量 $(\eta_1,\cdots,\eta_n)$ 序列在时刻 i 的值, 那么上面建立的正交随机变量 $\beta_1,\cdots,\beta_n$ 系的值 β_i, 不仅依赖于 "过去" $(\eta_1,\cdots,\eta_i)$, 而且依赖于 "将来" $(\eta_{i+1},\cdots,\eta_n)$. 下面将要介绍的**克拉默 – 施密特** (E. Schmidt) **正交化过程**不但没有这一缺陷, 而且它具有如下优点: 它可以用于**线性无关的**无限随机变量序列 (即任何有限个随机变量序列都线性无关).

设 $\eta_1,\eta_2,\cdots$ 是 L^2 中线性无关的随机变量序列. 用归纳法以下面介绍的方式建立序列 $\varepsilon_1,\varepsilon_2,\cdots$. 设 $\varepsilon_1=\eta_1/\|\eta_1\|$. 假设 $\varepsilon_1,\cdots,\varepsilon_{n-1}$ 已经选定, 而且它们正交, 则设

$$\varepsilon_n=\frac{\eta_n-\widehat{\eta}_n}{\|\eta_n-\widehat{\eta}_n\|}, \tag{11}$$

其中 $\widehat{\eta}_n$ 是 η_n 在由正交随机变量 $\varepsilon_1,\cdots,\varepsilon_{n-1}$ 生成的线性流形 $\mathscr{L}=\mathscr{L}(\varepsilon_1,\cdots,\varepsilon_{n-1})$ 上的投影:

$$\widehat{\eta}_n=\sum_{k=1}^{n-1}(\eta_n,\varepsilon_k)\varepsilon_k. \tag{12}$$

由于 $\eta_1,\cdots,\eta_n$ 线性无关, $\mathscr{L}(\eta_1,\cdots,\eta_{n-1})=\mathscr{L}(\varepsilon_1,\cdots,\varepsilon_{n-1})$ 可见 $\|\eta_n-\widehat{\eta}_n\|>0$, 从而 ε_n 有定义.

根据构造 $\|\varepsilon_n\|=1, n\geqslant 1$, 而且显然 $(\varepsilon_n,\varepsilon_k)=0, k<n$. 因此 $\varepsilon_1,\varepsilon_2,\cdots$ 是正交序列. 这时根据 (11) 式有

$$\eta_n=\widehat{\eta}_n+b_n\varepsilon_n,$$

其中 $b_n=\|\eta_n-\widehat{\eta}_n\|$, 而 $\widehat{\eta}_n$ 决定于 (12) 式.

现在假设 $\eta_1,\cdots,\eta_n$ 是任意随机变量系 (未必是线性无关的). 设 $\det \boldsymbol{R}=0$, 其中 $\boldsymbol{R}\equiv(r_{ij})$ 是向量 $(\eta_1,\cdots,\eta_n)$ 的协方差矩阵; 设

$$\operatorname{rang}\boldsymbol{R}=r<n.$$

那么, 由代数熟知, 对于二次型

$$Q(a)=\sum_{i,j=1}^{n}r_{ij}a_ia_j,\quad a=(a_1,\cdots,a_n),$$

恰好存在 $n-r$ 个线性无关向量 $a^{(1)},\cdots,a^{(n-r)}$, 使 $Q(a^{(i)})=0, i=1,\cdots,n-r$.

因为

$$Q(a)=\mathrm{E}\left(\sum_{k=1}^{n}a_k\eta_k\right)^2,$$

从而, 依概率 1, 有

$$\sum_{k=1}^{n}a_k^{(i)}\eta_k=0, i=1,\cdots,n-r.$$

换句话说, 在随机变量 $\eta_1, \cdots, \eta_n$ 之间, 恰好存在 $n-r$ 个线性关系式. 因此, 例如 $\eta_1, \cdots, \eta_r$ 线性无关, 则其余一切随机变量 $\eta_{r+1}, \cdots, \eta_n$ 可以通过 $\eta_1, \cdots, \eta_r$ 线性表示, 故 $\mathscr{L}(\eta_1, \cdots, \eta_n) = \mathscr{L}(\eta_1, \cdots, \eta_r)$. 由此可见, 利用正交化过程, 可以得到 r 个正交随机变量 $\varepsilon_1, \cdots, \varepsilon_r$, 使之一切通过 $\eta_1, \cdots, \eta_n$ 线性表示, 并且 $\mathscr{L}(\eta_1, \cdots, \eta_n) = \mathscr{L}(\varepsilon_1, \cdots, \varepsilon_r)$.

5. 正交基底和正交化 假设 $\eta_1, \eta_2, \cdots$ 是 L^2 中的随机变量序列. 以 $\mathscr{L}(\eta_1, \cdots, \eta_n)$ 表示随机变量 $\eta_1, \eta_2, \cdots$ 生成的*线性流形*, 即形如 $\sum_{i=1}^n a_i\eta_i, n \geqslant 1, a_i \in \mathbb{R}$ 的随机变量的全体. 以 $\overline{\mathscr{L}} = \overline{\mathscr{L}}(\eta_1, \cdots, \eta_n)$ 表示随机变量 $\eta_1, \eta_2, \cdots$ 生成的*闭线性流形*, 即 $\mathscr{L}$ 中的随机变量 $\eta_1, \eta_2, \cdots$ 及其均方极限的全体.

称随机变量系 $\eta_1, \eta_2, \cdots$ 是空间 L^2 中*可数正交基* (或*完全正交系*), 如果:

a) $\eta_1, \eta_2, \cdots$ 是正交系,

b) $\overline{\mathscr{L}}(\eta_1, \eta_2, \cdots) = L^2$.

具有可数正交基的希尔伯特空间, 称做*可分的*.

由于条件 b), 对于任意 $\xi \in L^2$ 和 $\varepsilon > 0$, 存在实数 $a_1, \cdots, a_n$, 使

$$\left\|\xi - \sum_{i=1}^n a_i\eta_i\right\| \leqslant \varepsilon.$$

那么, 根据 (3) 式,

$$\left\|\xi - \sum_{i=1}^n (\xi, \eta_i)\eta_i\right\| \leqslant \varepsilon$$

从而, 对于可分正交基的希尔伯特空间 L^2, 任意随机元素 ξ 可以表示为

$$\xi = \sum_{i=1}^\infty (\xi, \eta_i)\eta_i, \tag{13}$$

确切地说,

$$\xi = \underset{n}{\text{l.i.m.}} \sum_{i=1}^n (\xi, \eta_i)\eta_i.$$

由此及 (3) 式, 可见有如下*帕塞瓦尔* (M. A. Parseval) *等式*:

$$\|\xi\|^2 = \sum_{i=1}^\infty |(\xi, \eta_i)|^2, \ \xi \in L^2. \tag{14}$$

不难证明其逆命题也成立: 若 $\eta_1, \eta_2, \cdots$ 是某一正交系, 且 (13) 式或 (14) 式中任何条件成立, 则此正交系是基底.

举几个可分希尔伯特空间及其基底的例子.

例 1 设 $\Omega = \mathbb{R}, \mathscr{F} = \mathscr{B}(\mathbb{R})$, 而 $\mathbf{P}$ 是高斯测度:

$$\mathbf{P}(-\infty, a] = \int_{-\infty}^a \varphi(x)\mathrm{d}x, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}.$$

记 $D = \mathrm{d}/\mathrm{d}x$ 并引进函数

$$H_n(x) = \frac{(-1)^n D^n \varphi(x)}{\varphi(x)}, \quad n \geqslant 0. \tag{15}$$

不难求出:

$$\begin{aligned} D\varphi(x) &= -x\varphi(x), \\ D^2\varphi(x) &= (x^2-1)\varphi(x), \\ D^3\varphi(x) &= (3x-x^3)\varphi(x), \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{16}$$

由此可见, $H_n(x)$ 是多项式 (称做埃尔米特 [G. Hermite] 多项式). 由 (15) 和 (16) 式, 有

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2-1, \\ H_3(x) &= x^3-3x, \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

由简单的计算, 可得

$$(H_m, H_n) = \int_{-\infty}^{\infty} H_m(x)H_n(x)\mathbf{P}(\mathrm{d}x) = \int_{-\infty}^{\infty} H_m(x)H_n(x)\varphi(x)\mathrm{d}x = n!\delta_{nm},$$

其中 δ_{nm} 是克罗内克 (L. Kronecker) 记号 ($\delta_{nm}=0$, 若 $m \neq n$; $\delta_{nm}=1$, 若 $m=n$). 因此, 如果设

$$h_n(x) = \frac{H_n(x)}{\sqrt{n!}},$$

则这些赋范埃尔米特多项式 $\{h_n(x)\}_{n\geqslant 0}$ 为规范正交基. 由泛函分析 [33, 第七章 §3] 知, 如果

$$\lim_{c\downarrow 0} \int_{-\infty}^{\infty} \mathrm{e}^{c|x|}\mathbf{P}(\mathrm{d}x) < \infty, \tag{17}$$

则函数系 $\{1, x, x^2, \cdots\}$ 在 L^2 中是完备的, 即 L^2 中的任意函数 $\xi = \xi(x)$, 要么可以表示为 $\sum_{i=1}^n a_i\eta_i(x)$, 其中 $\eta_i(x) = x^i$; 要么可以表示为这些和 (在均方意义上) 的极限. 如果对于序列 $\eta_1(x), \eta_2(x), \cdots (\eta_i(x) = x^i)$ 运用克拉默 – 施密特正交化过程, 则所得规范化正交系恰好与赋范埃尔米特多项式系相同. 条件 (17) 对于现在考虑的情形成立. 从而, 多项式 $\{h_n(x)\}_{n\geqslant 0}$ 是基底, 说明在所考虑的概率空间中, 任意随机变量 $\xi = \xi(x)$ 都可以表示为

$$\xi(x) = \underset{n}{\mathrm{l.i.m.}} \sum_{i=1}^n (\xi, h_i)h_i(x). \tag{18}$$

例 2 设 $\Omega=\{0,1,2,\cdots\}, P=\{P_1,P_2,\cdots\}$ 是泊松分布:

$$P_x=\frac{\lambda^x}{x!}\mathrm{e}^{-\lambda}, x=0,1,\cdots;\lambda>0.$$

记 $\Delta f(x)=f(x)-f(x-1)(f(x)=0, x<0)$, 并仿照 (15) 式定义泊松 – 沙利耶 (M. Charlier) 多项式

$$\Pi_n(x)=\frac{(-1)^n\Delta^n P_x}{P_x},\quad n\geqslant 1, \Pi_0=1. \tag{19}$$

由于

$$(\Pi_m,\Pi_n)=\sum_{x=0}^{\infty}\Pi_m(x)\Pi_n(x)P_x=c_n\delta_{mn},$$

其中 c_n 是正常数, 则赋范泊松–沙利耶多项式 $\{\pi_n(x)\}_{n\geqslant 0}$, 其中 $\pi_n(x)=\Pi_n(x)/\sqrt{c_n}$ 为规范正交系, 并且因为满足条件 (17) 式, $\{\pi_n(x)\}_{n\geqslant 0}$ 也是规范化正交基.

例 3 在本例中将要引进的拉德马赫 – 哈尔 (H. Rademacher–A. Haar) 规范正交函数系, 不但在函数论中, 而且在概率论中也很重要.

设 $\Omega=[0,1), \mathscr{F}=\mathscr{B}([0,1))$, 而 $\mathbf{P}$ 是勒贝格测度. 在 §1 曾经指出, 每一个数 $x\in[0,1)$ 都可以表示为二进制小数:

$$x=\frac{x_1}{2}+\frac{x_2}{2^2}+\cdots,$$

其中 $x_i=0$ 或 1. (为了唯一性, 我们约定: 只考虑在二进制小数中含无限个 0 的数. 例如, 对于

$$\frac{1}{2}=\frac{1}{2}+\frac{0}{2^2}+\frac{0}{2^3}+\cdots=\frac{0}{2}+\frac{1}{2^2}+\frac{1}{2^3}+\cdots,$$

我们只取第一个.)

引进随机变量 $\xi_1(x),\xi_2(x),\cdots$, 设

$$\xi_n(x)=x_n.$$

那么, 对于只有 0 和 1 两个可能值的任何 a_i, 有

$$\begin{aligned}
&\mathbf{P}\{x:\xi_1=a_1,\cdots,\xi_n=a_n\}\\
&=\mathbf{P}\left\{x:\frac{a_1}{2}+\frac{a_2}{2^2}+\cdots+\frac{a_n}{2^n}\leqslant x<\frac{a_1}{2}+\frac{a_2}{2^2}+\cdots+\frac{a_n}{2^n}+\frac{1}{2^n}\right\}\\
&=\mathbf{P}\left\{x:x\in\left[\frac{a_1}{2}+\frac{a_2}{2^2}+\cdots+\frac{a_n}{2^n},\frac{a_1}{2}+\frac{a_2}{2^2}+\cdots+\frac{a_n}{2^n}+\frac{1}{2^n}\right)\right\}=\frac{1}{2^n}.
\end{aligned}$$

由此可见, $\xi_1,\xi_2,\cdots$ 是独立伯努利随机变量序列 (图 30 表示 $\xi_1=\xi_1(x)$ 和 $\xi_2=\xi_2(x)$ 的构造).

现在如果设 $R_n(x)=1-2\xi_n(x), n\geqslant 1$, 则不难验证函数系 $\{R_n\}$ (图 31 是拉德马赫函数) 是正交的:

$$\mathbf{E}R_nR_m=\int_0^1 R_n(x)R_m(x)\mathrm{d}x=\delta_{nm}.$$

注意 $(1, R_n) \equiv \mathbf{E}R_n = 0$. 由此可见该函数系不是完备的.

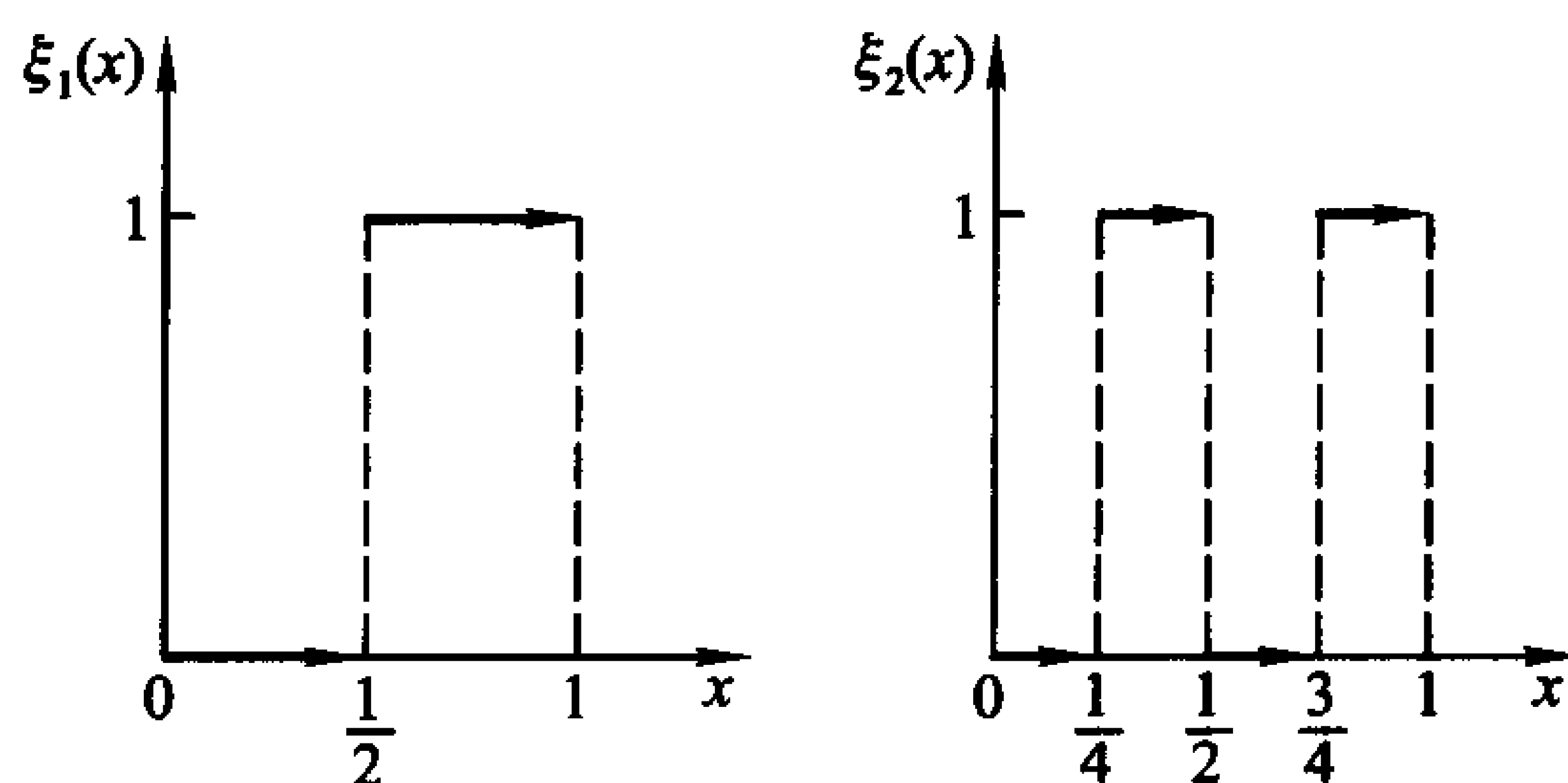


图 30 伯努利变量

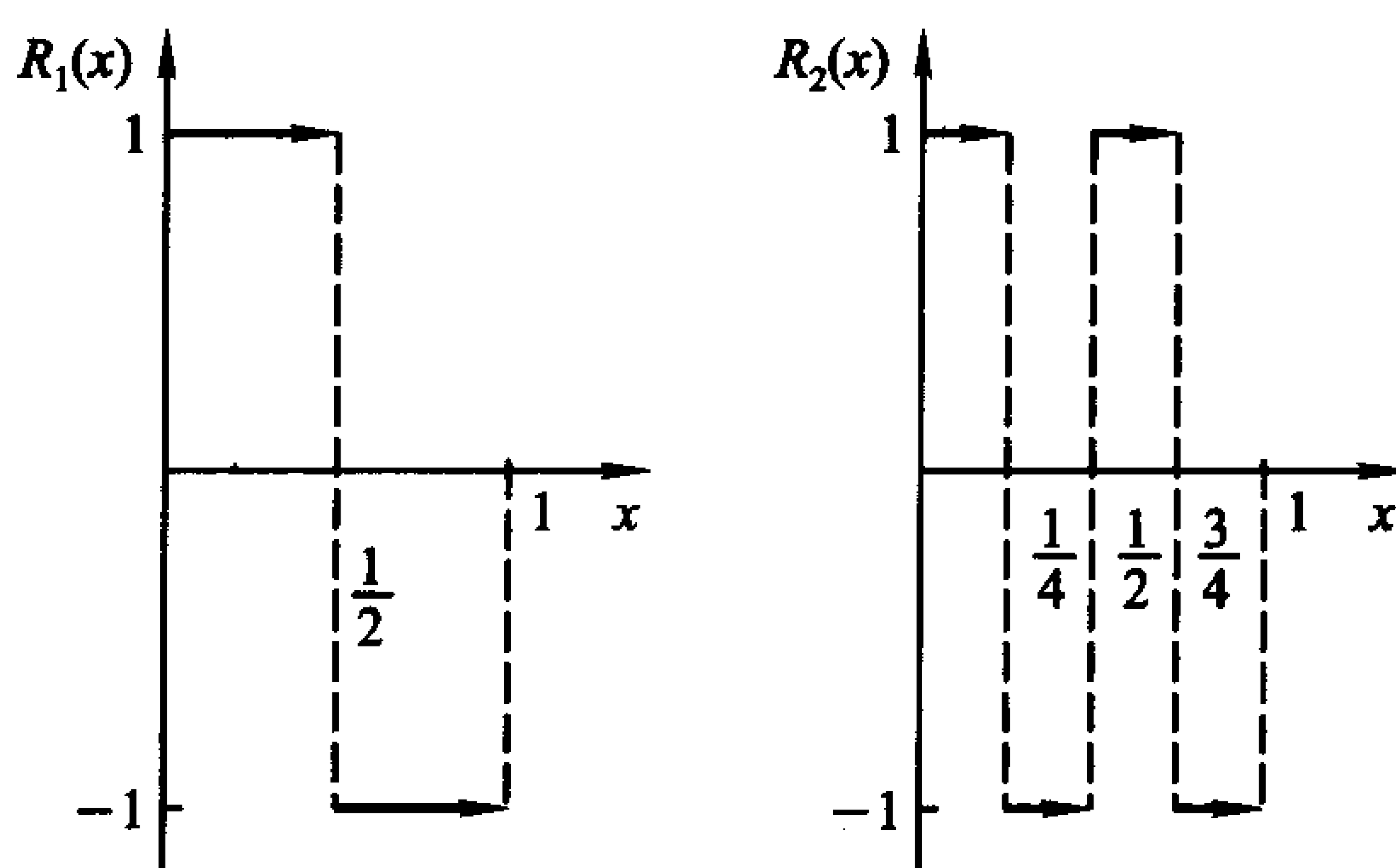


图 31 拉德马赫函数

然而, 拉德马赫系可以用来构造所谓哈尔系, 而哈尔系便于操作, 且不仅是规范正交的, 并且是完备的.

仍然假设 $\Omega = [0,1), \mathscr{F} = \mathscr{B}([0,1))$. 设

$$H_1(x) = 1,$$

$$H_2(x) = R_1(x),$$

$$\cdots\cdots\cdots\cdots$$

$$H_n(x) = \begin{cases} 2^{j/2} R_{j+1}(x), & \text{若 } \dfrac{k-1}{2^j} \leqslant x < \dfrac{k}{2^j}, n = 2^j + k, 1 \leqslant k \leqslant 2^j, j \geqslant 1, \\ 0, & \text{若不然}. \end{cases}$$

不难验证, $H_n(x), n \geqslant 3$, 可以表示为:

$$H_{2^m+1}(x) = \begin{cases} 2^{m/2}, & \text{若 } 0 \leqslant x < 2^{-(m+1)}, \\ -2^{m/2}, & \text{若 } 2^{-(m+1)} \leqslant x < 2^{-m}, \\ 0, & \text{其他}. \end{cases}$$

$$H_{2^m+j}(x) = H_{2^m+1}\left(x - \frac{j-1}{2^m}\right), j = 1, \cdots, 2^m, m = 1, 2, \cdots.$$

图 32 是前 8 个哈尔函数的示意图. 由图 32 可以得到哈尔函数组成的构造和性质的印象.

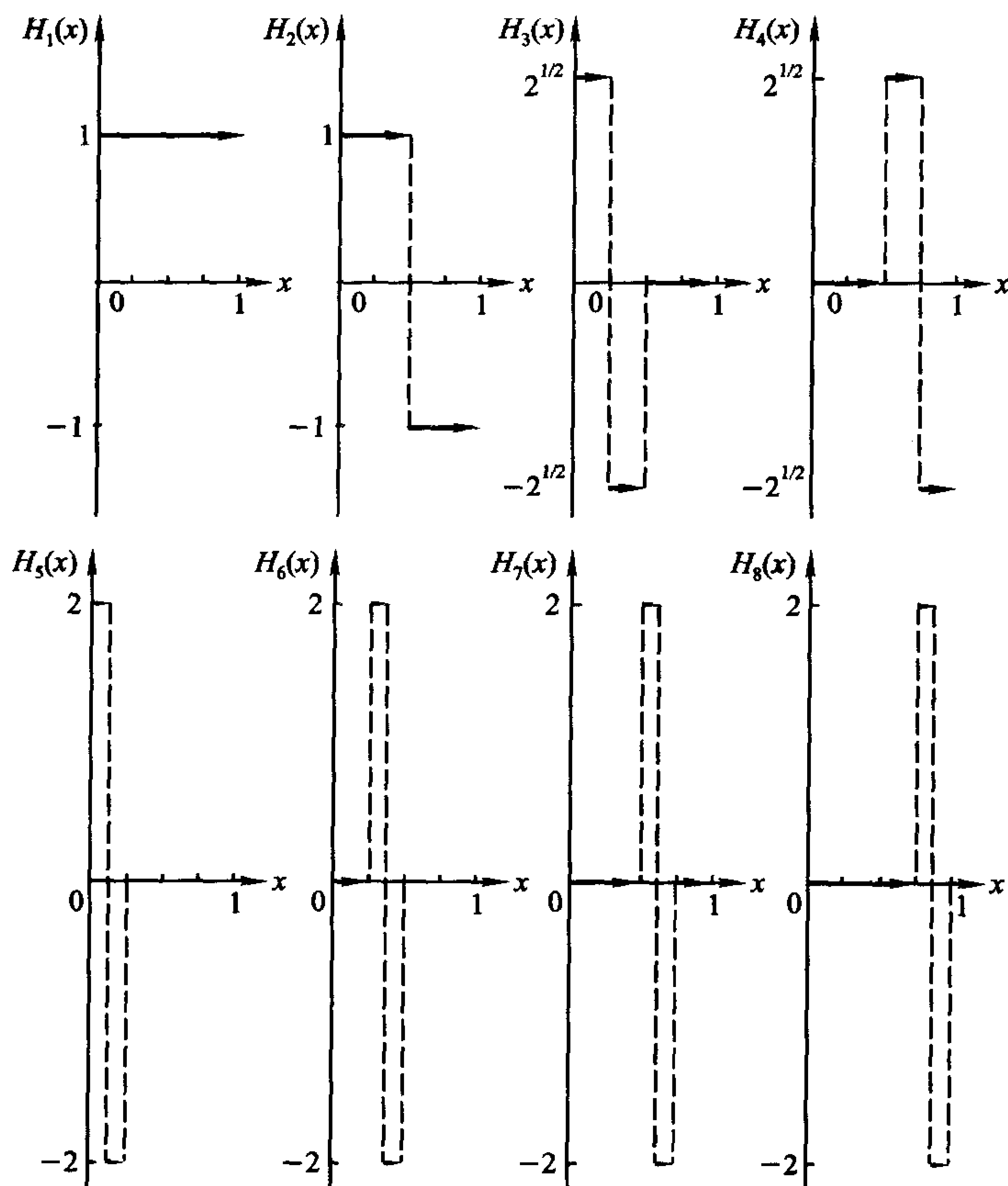


图 32 哈尔函数 $H_1(x),\cdots,H_8(x)$

不难验证, 哈尔函数系是规范化正交的, 并且在 L^1 和 L^2 中是完备的, 即如果对于 $p=1$ 或 2, 哈尔函数 $f=f(x)\in L^p$, 则

$$\int_0^1\left|f(x)-\sum_{k=1}^{n}(f,H_k)H_k(x)\right|^p \mathrm{d}x\to 0, n\to\infty;$$

而且, 依概率 1 (对勒贝格测度) 有

$$\sum_{k=1}^{n}(f,H_k)H_k(x)\to f(x), n\to\infty.$$

我们将在第七章 §4 证明这些事实. 这些结果由鞅收敛的一般理论导出, 并且是鞅方法在函数论中的运用很好的实例.

6. 最优线性估计 如果 $\eta_1,\cdots,\eta_n$ 是某一有限规范正交系, 则如上面已经证明的, 对于任意随机变量 $\xi\in L^2$, 在线性流形 $\mathscr{L}=\mathscr{L}(\eta_1,\cdots,\eta_n)$ 中, 存在随机变量 $\widehat{\xi}$(ξ 在 $\mathscr{L}$ 上的投影), 使

$$\|\xi-\widehat{\xi}\|=\inf_{\zeta}\{\|\xi-\zeta\|:\zeta\in\mathscr{L}(\eta_1,\cdots,\eta_n)\}.$$

这时 $\widehat{\xi}=\sum_{i=1}^n(\xi,\eta_i)\eta_i$. 这一结果容许自然地推广到 $\eta_1,\eta_2,\cdots$ 是可数正交系 (未必是基底) 的情形. 具体地说, 有如下结果.

定理 *假设 $\eta_1,\eta_2,\cdots$ 是规范正交随机变量系, $\overline{\mathscr{L}}=\overline{\mathscr{L}}(\eta_1,\eta_2,\cdots)$ 是由 $\eta_1,\eta_2,\cdots$ 生成的封闭线性流形. 那么, 存在并且唯一随机元 $\widehat{\xi}\in\overline{\mathscr{L}}$, 使*

$$\|\xi-\widehat{\xi}\|=\inf_{\zeta}\{\|\xi-\zeta\|:\zeta\in\overline{\mathscr{L}}\}.\tag{20}$$

这时,

$$\widehat{\xi}=\underset{n}{\mathrm{l.i.m.}}\sum_{i=1}^n(\xi,\eta_i)\eta_i,\tag{21}$$

且 $\xi-\widehat{\xi}\perp\zeta,\zeta\in\overline{\mathscr{L}}$.

证明 记 $d=\inf\{\|\xi-\zeta\|:\zeta\in\overline{\mathscr{L}}\}$, 并选一序列 $\zeta_1,\zeta_2,\cdots$, 使 $\|\xi-\zeta_n\|\to d$. 现在证明此序列是基本的. 简单的计算表明,

$$\|\zeta_n-\zeta_m\|^2=2\|\zeta_n-\xi\|^2+2\|\zeta_m-\xi\|^2-4\left\|\frac{1}{2}(\zeta_n+\xi_m)-\xi\right\|^2.$$

由于明显 $(\zeta_n+\xi_m)/2\in\overline{\mathscr{L}}$, 可见

$$\left\|\frac{1}{2}(\zeta_n+\zeta_m)-\xi\right\|^2\geqslant d^2,$$

从而 $\|\zeta_n-\zeta_m\|^2\to 0, n,m\to\infty$.

空间 L^2 是完备的 (§10 定理 7). 因此存在随机元 $\widehat{\xi}$, 使 $\|\zeta_n-\widehat{\xi}\|\to 0$. 由于集合 $\overline{\mathscr{L}}$ 是封闭的, 故 $\widehat{\xi}\in\overline{\mathscr{L}}$. 其次, 因为 $\|\xi-\zeta_n\|\to d$, 所以 $\|\xi-\widehat{\xi}\|=d$; 从而证明了所需要元素 $\widehat{\xi}$ 的存在性.

现在证明 $\widehat{\xi}$ 是 $\overline{\mathscr{L}}$ 中唯一具有所要求性质的元素. 假设 $\overline{\xi}\in\overline{\mathscr{L}}$ 且

$$\|\xi-\widehat{\xi}\|=\|\xi-\overline{\xi}\|=d.$$

那么 (由于练习题 3)

$$\|\widehat{\xi}+\overline{\xi}-2\xi\|^2+\|\widehat{\xi}-\overline{\xi}\|^2=2\|\widehat{\xi}-\xi\|^2+2\|\overline{\xi}-\xi\|^2=4d^2.$$

但是

$$\|\widehat{\xi}+\overline{\xi}-2\xi\|^2=4\left\|\frac{1}{2}(\widehat{\xi}+\overline{\xi})-\xi\right\|^2\geqslant 4d^2.$$

从而, $\|\widehat{\xi}-\overline{\xi}\|^2=0$, 这就证明了在 $\overline{\mathscr{L}}$ 中 “最接近” ξ 的元素的唯一性.

现在证明 $\xi-\widehat{\xi}\perp\zeta,\zeta\in\overline{\mathscr{L}}$. 由 (20) 式, 对于任意 $c\in\mathbb{R}$,

$$\|\xi-\widehat{\xi}-c\zeta\|\geqslant\|\xi-\widehat{\xi}\|.$$

由于

$$\|\xi-\widehat{\xi}-c\zeta\|^2=\|\xi-\widehat{\xi}\|^2+c^2\|\zeta\|^2-2(\xi-\widehat{\xi},c\zeta),$$

所以

$$c^2\|\zeta\|^2\geqslant 2(\xi-\widehat{\xi},c\zeta). \tag{22}$$

取 $c=\lambda(\xi-\widehat{\xi},\zeta),\lambda\in\mathbb{R}$, 则由 (22) 式, 得

$$(\xi-\widehat{\xi},\zeta)^2[\lambda^2\|\zeta\|^2-2\lambda]\geqslant 0.$$

对于充分小的正数 λ, 不等式 $\lambda^2\|\zeta\|^2-2\lambda<0$ 成立. 因此 $(\xi-\widehat{\xi},\zeta)=0,\zeta\in\overline{\mathscr{L}}$.

只剩下证明 (21) 式.

集合 $\overline{\mathscr{L}}=\overline{\mathscr{L}}(\eta_1,\eta_2,\cdots)$ 是 L^2 的封闭子空间, 因此本身是希尔伯特空间 (具有与 L^2 同样的数量积). 随机变量族 $\eta_1,\eta_2,\cdots$ 是希尔伯特空间 $\overline{\mathscr{L}}$ 的基底, 从而

$$\widehat{\xi}=\underset{n}{\mathrm{l.i.m.}}\sum_{k=1}^{n}(\xi,\eta_k)\eta_k. \tag{23}$$

由于 $\xi-\widehat{\xi}\perp\eta_k,k\geqslant 1$, 可见 $(\widehat{\xi},\eta_k)=(\xi,\eta_k),k\geqslant 0$. 于是, 连同 (23) 式就证明了 (21) 式. □

注 像有限维情形一样, $\widehat{\xi}$ 称为在 $\overline{\mathscr{L}}=\overline{\mathscr{L}}(\eta_1,\eta_2,\cdots)$ 上的**投影**, $\xi-\widehat{\xi}$ 称做**垂线**, 而表达式

$$\xi=\widehat{\xi}+(\xi-\widehat{\xi})$$

称做**正交分解**.

随机变量 $\widehat{\xi}$ 记作 $\widehat{\mathbf{E}}(\xi|\eta_1,\eta_2,\cdots)$ (对照第 3 小节的 $\widehat{\mathbf{E}}(\xi|\eta_1,\cdots,\eta_n)$), 并且称做 ($\xi$ 关于 $\eta_1,\eta_2,\cdots$ 的) **广义条件数学期望**. 按由 $\eta_1,\eta_2,\cdots$ 估计 ξ 的观点, 随机变量 $\widehat{\xi}$ 是**最优线性估计量**, 而由 (5) 和 (23) 式可见, 估计量的误差为

$$\Delta\equiv\mathbf{E}|\xi-\widehat{\xi}|^2\equiv\|\xi-\widehat{\xi}\|^2=\|\xi\|^2-\sum_{i=1}^{\infty}|(\xi,\eta_i)|^2.$$

7. 练习题

1. 证明, 如果 $\xi=\mathrm{l.i.m.}\xi_n$, 则 $\|\xi_n\|\to\|\xi\|$.
2. 证明, 如果 $\xi=\mathrm{l.i.m.}\xi_n$ 和 $\eta=\mathrm{l.i.m.}\eta_n$, 则 $(\xi_n,\eta_n)\to(\xi,\eta)$.
3. 证明, 范数 $\|\cdot\|$ 满足 “平行四边形” 性质:

$$\|\xi+\eta\|^2+\|\xi-\eta\|^2=2(\|\xi\|^2+\|\eta\|^2).$$

4. 证明, 正交随机变量族 $\{\xi_1, \cdots, \xi_n\}$, 满足 "毕达哥拉斯 (Pythagoras) 定理":

$$\left\|\sum_{i=1}^{n} \xi_i\right\|^2 = \sum_{i=1}^{n} \|\xi_i\|^2.$$

5. 设 $\xi_1, \xi_2, \cdots$ 是正交随机变量序列, $S_n = \xi_1 + \cdots + \xi_n$. 证明, 若 $\sum_{n=1}^{\infty} \mathbf{E}\xi_n^2 < \infty$, 则存在随机变量 $S(\mathbf{E}S^2 < \infty)$, 使 $\text{l.i.m.} S_n = S$, 即 $\|S_n - S\|^2 = \mathbf{E}|S_n - S|^2 \to 0, n \to \infty$.

6. 证明拉德马赫函数 R_n 可以定义为:

$$R_n(x) = \text{sign}(\sin 2^n \pi x), \quad 0 \leqslant x \leqslant 1, \quad n = 1, 2, \cdots$$

7. 证明, 对于 $\xi \in L^2(\mathscr{F})$,

$$\|\xi\| \geqslant \|\mathbf{E}(\xi|\mathscr{G})\|,$$

并且等式成立, 当且仅当几乎必然 $\xi = \mathbf{E}(\xi|\mathscr{G})$.

8. 证明, 若 $\xi, \eta \in L^2(\mathscr{F}), \mathbf{E}(\xi|\eta) = \eta, \mathbf{E}(\eta|\xi) = \xi$, 则 $\mathbf{P}\{\xi = \eta\} = 1$.

9. 有 $\mathscr{F}$ 的 3 个子 $\sigma-$ 代数序列: $(\mathscr{G}_n^{(1)}), (\mathscr{G}_n^{(2)})$ 和 $(\mathscr{G}_n^{(3)})$. 设 ξ 是有界随机变量. 已知对于每个 n, 有

$$\mathscr{G}_n^{(1)} \subseteq \mathscr{G}_n^{(2)} \subseteq \mathscr{G}_n^{(3)}, \quad \mathbf{E}(\xi|\mathscr{G}_n^{(1)}) \xrightarrow{\mathbf{P}} \eta, \quad \mathbf{E}(\xi|\mathscr{G}_n^{(3)}) \xrightarrow{\mathbf{P}} \eta.$$

证明 $\mathbf{E}(\xi|\mathscr{G}_n^{(2)}) \xrightarrow{\mathbf{P}} \eta$.

§12. 特征函数

1. 复数值随机变量 *特征函数方法*是概率论的基本的分析工具之一. 这在第三章中证明极限定理时, 特别是证明中心极限定理时表现得最明显, 中心极限定理是棣莫弗 - 拉普拉斯定理的推广. 这里, 我们将限于介绍特征函数的定义及其基本性质.

首先, 作一个一般性的说明.

除 (取实数值的) 随机变量外, 特征函数还要求考虑*复数值随机变量* (见 §5 第 1 小节).

随机变量的许多有关定义和性质, 都可以很容易移植到复数情形. 例如, 若数学期望为 $\mathbf{E}\xi$ 和 $\mathbf{E}\eta$, 则认为*复数值随机变量* $\zeta = \xi + i\eta$ 的数学期望定义为 $\mathbf{E}\zeta = \mathbf{E}\xi + i\mathbf{E}\eta$. 由随机元独立性的定义 6 (§5) 不难得到, 复数值随机变量 $\zeta_1 = \xi_1 + i\eta_1$ 和 $\zeta_2 = \xi_2 + i\eta_2$ 相互独立, 当且仅当随机向量 (ξ_1, η_1) 和 (ξ_2, η_2) 独立, 或等价地 $\sigma-$ 代数 $\mathscr{F}_{\xi_1, \eta_1}$ 和 $\mathscr{F}_{\xi_2, \eta_2}$ 相互独立.

与具有有限二阶矩的实数值随机变量的空间 L^2 同时, 我们引进复数值随机变量 $\zeta = \xi + i\eta$ 的*希尔伯特*空间, 其中 $\mathbf{E}|\zeta| < \infty, |\zeta|^2 = \xi^2 + \eta^2$, 并在其中引进数量积

$(\zeta_1,\zeta_2)=\mathbf{E}\zeta_1\overline{\zeta}_2$, 其中 $\overline{\zeta}_2$ 是复值共轭随机变量. 以后实数值随机变量和复数值随机变量, 一般都称为随机变量, 除非必要时才明确指出具体是哪一种.

还需要指出下列记号. 在向量 $a\in\mathbb{R}^n$ 的代数运算中, 向量视为列向量:

$$a=\begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix},$$

并以 $a^{\mathrm{T}}=(a_1,\cdots,a_n)$ 表示行向量. 对于 $a,b\in\mathbb{R}^n$, 把 $\sum_{i=1}^n a_ib_i$ 理解为 a 和 b 的数量积 (a,b). 显然, $(a,b)=a^{\mathrm{T}}b$.

如果 $a\in\mathbb{R}^n$, 而 $\boldsymbol{R}=(r_{ij})$ 是 $n\times n$ 阶矩阵, 则

$$(\boldsymbol{R}a,a)=a^{\mathrm{T}}\boldsymbol{R}a=\sum_{i,j=1}^n a_ir_{ij}a_j. \tag{1}$$

2. 特征函数的定义

定义 1 设 $F=F(\boldsymbol{x}),\boldsymbol{x}=(x_1,\cdots,x_n)$ 是 $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ 上的 n 维分布函数, 则函数

$$\varphi(t)=\int_{\mathbb{R}^n}\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{x})}\mathrm{d}F(\boldsymbol{x}),\quad t\in\mathbb{R}^n, \tag{2}$$

称做 $F(\boldsymbol{x})$ 的特征函数.

定义 2 设 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 是空间 $(\Omega,\mathscr{F},\mathbf{P})$ 上在 $\mathbb{R}^n$ 中取值的随机向量, 则函数

$$\varphi_\xi(\boldsymbol{t})=\int_{\mathbb{R}^n}\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{x})}\mathrm{d}F_\xi(\boldsymbol{x}),\quad \boldsymbol{t}\in\mathbb{R}^n, \tag{3}$$

称做 $F_\xi(\boldsymbol{x})$ 的特征函数, 其中 $F_\xi=F_\xi(\boldsymbol{x}),\boldsymbol{x}=(x_1,\cdots,x_n)$ 是随机向量 $\xi=(\xi_1,\cdots,\xi_n)$ 的分布函数.

如果 $F(\boldsymbol{x})$ 有密度 $f=f(\boldsymbol{x})$, 则

$$\varphi(\boldsymbol{t})=\int_{\mathbb{R}^n}\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{x})}f(\boldsymbol{x})\mathrm{d}\boldsymbol{x},\quad \boldsymbol{t}\in\mathbb{R}^n.$$

换句话说, 这时特征函数 $\varphi(\boldsymbol{t})$ 恰好是函数 $f=f(\boldsymbol{x})$ 的傅里叶 (J. Fourier) 变换.

由 (3) 式和 §6 (关于在勒贝格积分号下求极限的) 定理 7, 可见随机向量的特征函数也可以由如下等式定义:

$$\varphi_\xi(\boldsymbol{t})=\mathbf{E}\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{\xi})},\quad \boldsymbol{t}\in\mathbb{R}^n, \tag{4}$$

现在讨论特征函数的基本性质, 但是只证明 $n=1$ 的情形. 一些涉及一般情形的相对重要的结果, 将在练习题中给出.

设 $\xi = \xi(\omega)$ 是随机变量, $F_\xi = F_\xi(x)$ 是其分布函数, 而

$$\varphi_\xi(t) = \mathbf{E}e^{it\xi}$$

是其特征函数.

立即可以指出, 若 $\eta = a\xi + b$, 则

$$\varphi_\eta(t) = \mathbf{E}e^{it\eta} = \mathbf{E}e^{it(a\xi+b)} = e^{itb}\mathbf{E}e^{ita\xi}.$$

因此

$$\varphi_\eta(t) = e^{itb}\varphi_\xi(at). \tag{5}$$

其次, 如果 $\xi_1, \xi_2, \cdots, \xi_n$ 是独立随机变量, 而 $S_n = \xi_1 + \cdots + \xi_n$, 则

$$\varphi_{S_n}(t) = \prod_{j=1}^n \varphi_{\xi_j}(t), \tag{6}$$

事实上, 由于 (有界) 独立随机变量乘积的数学期望 (无论是实数值随机变量, 还是复数值随机变量, 可见 §6 定理 6 和练习题 1), 等于其数学期望的乘积, 可见

$$\varphi_{S_n}(t) = \mathbf{E}e^{it(\xi_1+\cdots+\xi_n)} = \mathbf{E}e^{it\xi_1}\cdots\mathbf{E}e^{it\xi_n} = \prod_{i=1}^n \varphi_{\xi_i}(t).$$

对于用特征函数方法, 证明独立随机变量之和的极限定理 (第三章 §3), 性质 (6) 是关键, 这时, 通过和的各项表示分布函数 F_{S_n} 是相当复杂的, 具体地说,

$$F_{S_n} = F_{\xi_1} * \cdots * F_{\xi_n},$$

其中符号 “$*$” 表示分布的卷积 (见 §8 第 4 小节).

下面是特征函数的例子.

例 1　设ξ是伯努利随机变量, 且 $\mathbf{P}\{\xi = 1\} = p, \mathbf{P}\{\xi = 0\} = q, p + q = 1, 0 < p < 1$, 则

$$\varphi_\xi(t) = pe^{it} + q.$$

如果 $\xi_1, \cdots, \xi_n$ 是独立与 ξ 同分布的随机变量, 则对于 $T_n = (S_n - np)/\sqrt{npq}$, 有

$$\begin{aligned}\varphi_{T_n}(t) &= \varphi_{(S_n-np)/\sqrt{npq}}(t) = \mathbf{E}e^{it(S_n-np)/\sqrt{npq}} \\ &= e^{-it\sqrt{np/q}}[pe^{it/\sqrt{npq}} + q]^n = [pe^{it\sqrt{q/np}} + qe^{-it\sqrt{p/nq}}]^n.\end{aligned} \tag{7}$$

由此可见, 当 $n \to \infty$ 时, 有

$$\varphi_{T_n}(t) \to e^{-\frac{t^2}{2}}. \tag{8}$$

例 2 设 $\xi \sim N(m,\sigma^2), |m|<\infty, \sigma^2>0$ 现在证明

$$\varphi_\xi(t)=\mathrm{e}^{itm-\frac{t^2\sigma^2}{2}}. \tag{9}$$

设 $\eta=(\xi-m)/\sigma$, 则 $\eta \sim N(0,1)$, 而由 (5) 式

$$\varphi_\xi(t)=\mathrm{e}^{itm}\varphi_\eta(\sigma t),$$

故只需验证

$$\varphi_\eta(t)=\mathrm{e}^{-\frac{t^2}{2}}. \tag{10}$$

下面的一连串式子就可以证明 (10) 式:

$$\begin{aligned}\varphi_\eta(t)=\mathbf{E}\mathrm{e}^{it\eta}&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{itx}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sum_{n=0}^{\infty}\frac{(itx)^n}{n!}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\\&=\sum_{n=0}^{\infty}\frac{(it)^n}{n!}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^n\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x=\sum_{n=0}^{\infty}\frac{(it)^{2n}}{(2n)!}(2n-1)!!\\&=\sum_{n=0}^{\infty}\frac{(it)^{2n}}{(2n)!}\frac{(2n)!}{2^n n!}=\sum_{n=0}^{\infty}\left(-\frac{t^2}{2}\right)^n\frac{1}{n!}=\mathrm{e}^{-\frac{t^2}{2}},\end{aligned}$$

其中用到如下关系式 (见 §8 练习题 7):

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{2n}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\equiv\mathbf{E}\eta^{2n}=(2n-1)!!.$$

例 3 设 ξ 是泊松随机变量,

$$\mathbf{P}\{\xi=k\}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\quad k=0,1,\cdots$$

则

$$\mathbf{E}\mathrm{e}^{it\xi}=\sum_{k=0}^{\infty}\mathrm{e}^{itk}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\mathrm{e}^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda\mathrm{e}^{it})^k}{k!}=\exp\{\lambda(\mathrm{e}^{it}-1)\}. \tag{11}$$

3. 特征函数的性质 我们在 §9 第 1 小节曾指出, 对于 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上每一个分布函数 $F(x)$, 都存在一随机变量 ξ, 使 $F(x)$ 恰好是 ξ 分布函数. 因此, 在叙述特征函数时 (无论是按定义 1, 还是按定义 2), 可以局限于考虑随机变量 $\xi=\xi(\omega)$ 的特征函数 $\varphi(t)=\varphi_\xi(t)$.

定理 1 设随机变量 ξ 的分布函数为 $F=F(x)$, 而

$$\varphi(t)=\mathbf{E}\mathrm{e}^{it\xi}$$

是其特征函数.

那么, 有下列性质:

1) $|\varphi(t)| \leqslant \varphi(0) = 1$;

2) $\varphi(t)$ 对于 $t \in \mathbb{R}$ 一致连续;

3) $\varphi(t) = \overline{\varphi(-t)}$;

4) $\varphi(t)$ 是实数值函数的充分和必要条件是, 其分布函数 $F = F(x)$ 对称, 即

$$\int_B \mathrm{d}F(x) = \int_{-B} \mathrm{d}F(x),\ B \in \mathscr{B}(\mathbb{R}),\ -B = \{-x : x \in B\};$$

5) 如果对于某个 $n \geqslant 1, \mathbf{E}|\xi|^n < \infty$, 则对于一切 $r \leqslant n$, 存在导数 $\varphi^{(r)}(t)$, 且

$$\varphi^{(r)}(t) = \int_{\mathbb{R}} (ix)^r \mathrm{e}^{itx} \mathrm{d}F(x), \tag{12}$$

$$\mathbf{E}\xi^r = \frac{\varphi^{(r)}(0)}{i^r}, \tag{13}$$

$$\varphi(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!} \mathbf{E}\xi^r + \frac{(it)^n}{n!} \varepsilon_n(t), \tag{14}$$

其中 $|\varepsilon_n(t)| \leqslant 3\mathbf{E}|\xi|^n$ 且 $\varepsilon_n(t) \to 0, t \to 0$;

6) 如果存在有限导数 $\varphi^{(2n)}(0)$, 则 $\mathbf{E}\xi^{2n} < \infty$;

7) 如果对于一切 $n \geqslant 1, \mathbf{E}|\xi|^n < \infty$ 且

$$\varlimsup_{n} \frac{(\mathbf{E}|\xi|^n)^{1/n}}{n} = \frac{1}{T} < \infty,$$

则对于一切 $|t| < T$,

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mathbf{E}\xi^n. \tag{15}$$

证明 性质 1) 和 3) 显然. 性质 2) 由下面的估计式

$$|\varphi(t+h) - \varphi(t)| = |\mathbf{E}\mathrm{e}^{it\xi}(\mathrm{e}^{ih\xi} - 1)| \leqslant \mathbf{E}|\mathrm{e}^{ih\xi} - 1|$$

和控制收敛定理, 以及当 $h \to 0$ 时 $\mathbf{E}|\mathrm{e}^{ih\xi} - 1| \to 0$, 可以得到.

证明性质 4). 假设分布函数 $F = F(x)$ 对称, 则对于有界博雷尔奇函数 $g(x)$, 有

$$\int_{\mathbb{R}} g(x) \mathrm{d}F(x) = 0$$

(注意, 对于简单奇函数, 这由函数 F 的对称性的定义立即可以得到). 因此

$$\int_{\mathbb{R}} \sin\ tx \mathrm{d}F(x) = 0,\ \text{故}\ \varphi(t) = \mathbf{E}\cos t\xi.$$

相反, 设 $\varphi(t)$ 是实数值函数, 则由于 3), 可见

$$\varphi_{-\xi}(t) = \varphi_\xi(-t) = \overline{\varphi_\xi(t)} = \varphi_\xi(t),\ t \in \mathbb{R}.$$

由此 (这将在下面的定理 2 证明), 可见随机变量 $-\xi$ 和 ξ 的分布函数 $F_{-\xi}$ 和 F_{ξ} 相等, 因此 (根据 §3 的定理 1), 对于 $B \in \mathscr{B}(\mathbb{R})$, 有

$$\mathbf{P}\{\xi \in B\} = \mathbf{P}\{-\xi \in B\} = \mathbf{P}\{\xi \in -B\}.$$

证明性质 5). 如果 $\mathbf{E}|\xi|^n < \infty$, 则由李亚普诺夫不等式 (§6 的 (28) 式), 有 $\mathbf{E}|\xi|^r < \infty, r \leqslant n$.

考虑关系式

$$\frac{\varphi(t+h)-\varphi(t)}{h} = \mathbf{E}\mathrm{e}^{it\xi}\left(\frac{\mathrm{e}^{ih\xi}-1}{h}\right).$$

由于

$$\left|\frac{\mathrm{e}^{ih\xi}-1}{h}\right| \leqslant |x| \quad 和 \quad \mathbf{E}|\xi| < \infty,$$

可见根据控制收敛定理, 存在

$$\lim_{h\to\infty} \mathbf{E}\mathrm{e}^{it\xi}\left(\frac{\mathrm{e}^{ih\xi}-1}{h}\right),$$

等于

$$\mathbf{E}\mathrm{e}^{it\xi}\lim_{h\to\infty}\left(\frac{\mathrm{e}^{ih\xi}-1}{h}\right) = i\mathbf{E}(\xi\mathrm{e}^{it\xi}) = i\int_{-\infty}^{\infty} x\mathrm{e}^{itx}\mathrm{d}F(x). \tag{16}$$

因此, 存在导数 $\varphi'(t)$ 且

$$\varphi'(t) = i\mathbf{E}(\xi\mathrm{e}^{it\xi}) = i\int_{-\infty}^{\infty} x\mathrm{e}^{itx}\mathrm{d}F(x).$$

用归纳法可证明, 导数 $\varphi^{(r)}(t)(1 < r \leqslant n)$ 的存在性, 和 (12) 式的正确性.

直接由 (12) 式得出 (13) 式. 现在证明表达式 (14).

由于对于实数 y, 有

$$\mathrm{e}^{iy} = \cos\ y + i\sin\ y = \sum_{k=0}^{n-1}\frac{(iy)^k}{k!} + \frac{(iy)^n}{n!}[\cos\theta_1 y + i\sin\theta_2 y],$$

其中 $|\theta_1| \leqslant 1, |\theta_2| \leqslant 1$, 可见

$$\mathrm{e}^{it\xi} = \sum_{k=0}^{n-1}\frac{(it\xi)^k}{k!} + \frac{(it\xi)^n}{n!}[\cos\theta_1(\omega)t\xi + i\sin\theta_2(\omega)t\xi]. \tag{17}$$

且

$$\mathbf{E}\mathrm{e}^{it\xi} = \sum_{k=0}^{n-1}\frac{(it)^k}{k!}\mathbf{E}\xi^k + \frac{(it)^n}{n!}[\mathbf{E}\xi^n + \varepsilon_n(t)], \tag{18}$$

其中

$$\varepsilon_n(t) = \mathbf{E}\{\xi^n[\cos\theta_1(\omega)t\xi + i\sin\theta_2(\omega)t\xi - 1]\}.$$

显然, $|\varepsilon_n(t)| \leqslant 3\mathbf{E}|\xi^n|$, 而且根据控制收敛定理, 当 $t \to 0$ 时, $\varepsilon_n(t) \to 0$.

证明性质 6). 用归纳法证明. 首先假设导数 $\varphi''(0)$ 存在并且有限. 证明 $\mathbf{E}\xi^2 < \infty$ 根据洛必达 (G. L'hospital) 法则和法图引理, 有

$$\begin{aligned}
\varphi''(0) &= \lim_{h\to 0} \frac{1}{2}\left[\frac{\varphi'(2h)-\varphi'(0)}{2h} + \frac{\varphi'(0)-\varphi'(-2h)}{2h}\right] \\
&= \lim_{h\to 0} \frac{2\varphi'(2h)-2\varphi'(-2h)}{8h} = \lim_{h\to 0} \frac{1}{4h^2}[\varphi(2h)-2\varphi(0)+\varphi(-2h)] \\
&= \lim_{h\to 0} \int_{-\infty}^{\infty} \left(\frac{\mathrm{e}^{ihx}-\mathrm{e}^{-ihx}}{2h}\right)^2 \mathrm{d}F(x) = -\lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{\sin hx}{hx}\right)^2 x^2\mathrm{d}F(x) \\
&\leqslant -\int_{-\infty}^{\infty} \lim_{h\to 0}\left(\frac{\sin hx}{hx}\right)^2 x^2\mathrm{d}F(x) = -\int_{-\infty}^{\infty} x^2\mathrm{d}F(x).
\end{aligned}$$

从而

$$\int_{-\infty}^{\infty} x^2\mathrm{d}F(x) \leqslant -\varphi''(0) < \infty.$$

现在假设 $\varphi^{(2k+2)}(0)$ 存在而且有限

$$\int_{-\infty}^{\infty} x^{2k}\mathrm{d}F(x) < \infty.$$

如果

$$\int_{-\infty}^{\infty} x^{2k}\mathrm{d}F(x) = 0, \quad 则 \quad \int_{-\infty}^{\infty} x^{2k+2}\mathrm{d}F(x) = 0.$$

因此, 我们将假设

$$\int_{-\infty}^{\infty} x^{2k}\mathrm{d}F(x) > 0.$$

那么, 根据性质 5), 有

$$\varphi^{(2k)}(t) = \int_{-\infty}^{\infty} (ix)^{2k}\mathrm{e}^{itx}\mathrm{d}F(x),$$

因而

$$(-1)^k\varphi^{(2k)}(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx}\mathrm{d}G(x),$$

其中 $G(x) = \int_{-\infty}^{x} u^{2k}\mathrm{d}F(u)$.

从而, 函数 $(-1)^k\varphi^{(2k)}(t)G^{-1}(\infty)$ 是概率分布 $G(x)\times G^{-1}(\infty)$ 的特征函数, 并且根据已证明的, 有

$$G^{-1}(\infty)\int_{-\infty}^{\infty} x^2\mathrm{d}G(x) < \infty.$$

由于 $G^{-1}(\infty) > 0$, 可见

$$\int_{-\infty}^{\infty} x^{2k+2}\mathrm{d}F(x) = \int_{-\infty}^{\infty} x^2\mathrm{d}G(x) < \infty.$$

证明性质 7). 设 $0 < t_0 < T$, 则利用斯特林公式 (第一章 §2 (6) 式), 有

$$\overline{\lim}\frac{\mathbf{E}(|\xi|^n)^{1/n}}{n} < \frac{1}{t_0} \Rightarrow \overline{\lim}\frac{\mathbf{E}(|\xi|^n t_0^n)^{1/n}}{n} < 1 \Rightarrow \overline{\lim}\left(\frac{\mathbf{E}|\xi|^n t_0^n}{n}\right)^{1/n} < 1.$$

从而, 根据柯西准则, 级数 $\sum \mathbf{E}|\xi|^n t_0^n/n!$ 收敛, 因此对于任意 $|t| \leqslant t_0$, 级数 $\sum_{r=0}^{\infty} \mathbf{E}\xi^r (it)^r/r!$ 收敛. 而由 (14) 式

$$\varphi(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!}\mathbf{E}\xi^r + R_n(t), n \geqslant 1,\ \text{其中}\ |R_n(t)| \leqslant \frac{3|t|^n}{n!}\mathbf{E}|\xi|^n.$$

于是, 对于一切 $|t| < T$, 有

$$\varphi(t) = \sum_{r=0}^{\infty} \frac{(it)^r}{r!}\mathbf{E}\xi^r. \qquad \square$$

注 1 与 (14) 式的证明类似, 可以验证, 如果对于某个 $n \geqslant 1, \mathbf{E}|\xi|^n < \infty$, 则

$$\varphi(t) = \sum_{k=0}^{n} \frac{i^k (t-s)^k}{k!} \int_{-\infty}^{\infty} x^k \mathrm{e}^{isx} \mathrm{d}F(x) + \frac{i^n (t-s)^n}{n!} \varepsilon_n(t-s), \tag{19}$$

其中 $|\varepsilon_n(t-s)| \leqslant 3\mathbf{E}|\xi^n|$, 而当 $t-s \to 0$ 时 $\varepsilon_n(t-s) \to 0$.

注 2 关于性质 7) 中的条件, 参见下面的第 9 小节中关于 "矩问题的唯一性" 的内容.

4. 特征函数唯一决定分布函数 下面的定理表明, 特征函数唯一决定分布函数.

定理 2 (唯一性) *假设 F 和 G 是具有同一特征函数的分布函数, 即对于一切 $t \in \mathbb{R}$,*

$$\int_{-\infty}^{\infty} \mathrm{e}^{itx} \mathrm{d}F(x) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} \mathrm{d}G(x). \tag{20}$$

那么 $F(x) = G(x)$.

证明 固定 $a, b \in \mathbb{R}, \varepsilon > 0$, 并且考虑图 33 所表示的函数 $f^\varepsilon = f^\varepsilon(x)$. 现在证明:

$$\int_{-\infty}^{\infty} f^\varepsilon(x) \mathrm{d}F(x) = \int_{-\infty}^{\infty} f^\varepsilon(x) \mathrm{d}G(x). \tag{21}$$

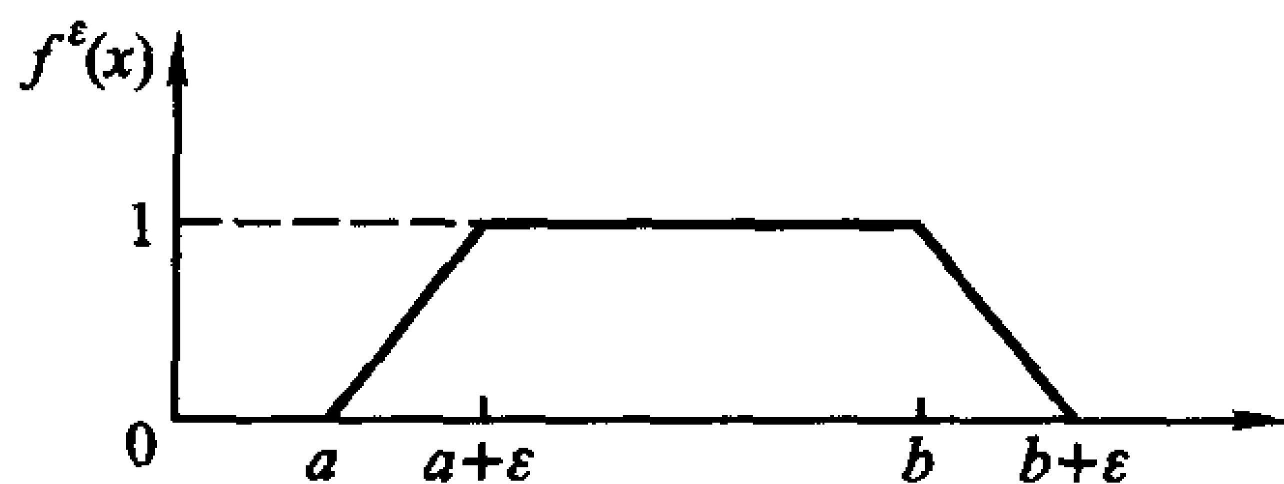


图 33

设 $n \geqslant 0$ 满足 $[a, b+\varepsilon] \subseteq [-n, n]$; 数列 $\{\delta_n\}$ 满足 $1 \geqslant \delta_n \downarrow 0, n \to \infty$. $f^\varepsilon = f^\varepsilon(x)$ 作为 $[-n, n]$ 上的连续且在端点上等值的函数, 可以用三角多项式一致逼近的 (维尔斯特拉斯 - 斯通 [M.Stone]) 定理, 即存在有限和

$$f_n^\varepsilon(x) = \sum_k a_k \exp\left(i\pi x \frac{k}{n}\right), \tag{22}$$

使

$$\sup_{-n \leqslant x \leqslant n} |f^\varepsilon(x) - f_n^\varepsilon(x)| \leqslant \delta_n. \tag{23}$$

对于所有 $x \in \mathbb{R}$ 将周期函数 $f_n^\varepsilon(x)$ 进行开拓, 并注意到

$$\sup_x |f_n^\varepsilon(x)| \leqslant 2.$$

那么, 由于 (20) 式

$$\int_{-\infty}^{\infty} f_n^\varepsilon(x) \mathrm{d}F(x) = \int_{-\infty}^{\infty} f_n^\varepsilon(x) \mathrm{d}G(x),$$

可见

$$\begin{aligned}
&\left|\int_{-\infty}^{\infty} f^\varepsilon(x) \mathrm{d}F(x) - \int_{-\infty}^{\infty} f^\varepsilon(x) \mathrm{d}G(x)\right| = \left|\int_{-n}^{n} f^\varepsilon \mathrm{d}F - \int_{-n}^{n} f^\varepsilon \mathrm{d}G\right| \\
\leqslant &\left|\int_{-n}^{n} f_n^\varepsilon \mathrm{d}F - \int_{-n}^{n} f_n^\varepsilon \mathrm{d}G\right| + 2\delta_n \\
\leqslant &\left|\int_{-\infty}^{\infty} f_n^\varepsilon \mathrm{d}F - \int_{-\infty}^{\infty} f_n^\varepsilon \mathrm{d}G\right| + 2\delta_n + 2F(\overline{[-n, n]}) + 2G(\overline{[-n, n]}),
\end{aligned} \tag{24}$$

其中

$$F(A) = \int_A \mathrm{d}F(x), \quad G(A) = \int_A \mathrm{d}G(x).$$

当 $n \to \infty$ 时 (24) 式的右侧趋向 0, 从而 (21) 式得证.

当 $\varepsilon \to 0$ 时 $f^\varepsilon(x) \to I_{(a,b]}(x)$. 因此根据控制收敛定理, 由 (21) 式, 可见

$$\int_{-\infty}^{\infty} I_{(a,b]}(x) \mathrm{d}F(x) = \int_{-\infty}^{\infty} I_{(a,b]}(x) \mathrm{d}G(x),$$

即 $F(b) - F(a) = G(b) - G(a)$, 而由于 a 和 b 是任意的, 可见对于一切 $x \in \mathbb{R}, F(x) = G(x)$. 于是, 定理 2 得证. □

5. 逆转公式 上一个定理说明, 分布函数 $F = F(x)$ 唯一决定于其特征函数 $\varphi = \varphi(t)$. 下面的定理将给出函数 F 通过函数 φ 的表达式.

定理 3 (逆转公式) 假设 $F = F(x)$ 是分布函数, 而

$$\varphi(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} \mathrm{d}F(x)$$

是其特征函数.

a) 对于函数 $F=F(x)$ 的任意两个连续点 a 和 $b(a<b)$, 有

$$F(b)-F(a)=\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\varphi(t)\mathrm{d}t. \tag{25}$$

b) 如果 $\int_{-\infty}^{\infty}|\varphi(t)|\mathrm{d}t<\infty$, 则分布函数 $F(x)$ 有密度 $f(x)$:

$$F(x)=\int_{-\infty}^{x}f(y)\mathrm{d}y, \tag{26}$$

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-itx}\varphi(t)\mathrm{d}t. \tag{27}$$

证明 首先注意到, 如果函数 $F(x)$ 有密度 $f(x)$, 则

$$\varphi(t)=\int_{-\infty}^{\infty}\mathrm{e}^{itx}f(x)\mathrm{d}x, \tag{28}$$

因此 (27) 式恰好是 (可积) 函数 $\varphi(x)$ 的傅里叶变换. 在 (27) 式两侧同时积分, 并运用傅比尼定理, 得

$$\begin{aligned}F(b)-F(a)&=\int_a^b f(x)\mathrm{d}x=\frac{1}{2\pi}\int_a^b\left[\int_{-\infty}^{\infty}\mathrm{e}^{-itx}\varphi(t)\mathrm{d}t\right]\mathrm{d}x\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)\left[\int_a^b\mathrm{e}^{-itx}\mathrm{d}x\right]\mathrm{d}t=\frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\mathrm{d}t.\end{aligned}$$

在完成了对 (25) 式一定的说明之后, 现在开始证明 (25) 式.

a) 有

$$\begin{aligned}\Phi_c&\equiv\frac{1}{2\pi}\int_{-c}^{c}\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\varphi(t)\mathrm{d}t=\frac{1}{2\pi}\int_{-c}^{c}\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\left[\int_{-\infty}^{\infty}\mathrm{e}^{itx}\mathrm{d}F(x)\right]\mathrm{d}t\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-c}^{c}\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\mathrm{e}^{itx}\mathrm{d}t\right]\mathrm{d}F(x)=\int_{-\infty}^{\infty}\Psi_c(x)\mathrm{d}F(x),\end{aligned} \tag{29}$$

其中, 设

$$\Psi_c(x)=\frac{1}{2\pi}\int_{-c}^{c}\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\mathrm{e}^{itx}\mathrm{d}t,$$

并且用到傅比尼定理: 由于

$$\left|\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\mathrm{e}^{itx}\right|=\left|\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\right|=\left|\int_a^b\mathrm{e}^{-itx}\mathrm{d}x\right|\leqslant b-a,$$

且

$$\int_{-c}^{c}\int_{-\infty}^{\infty}(b-a)\mathrm{d}F(x)\mathrm{d}t\leqslant 2c(b-a)<\infty,$$

可见应用傅比尼定理是合理的. 其次,

$$\begin{aligned}\Psi_c(x) &= \frac{1}{2\pi}\int_{-c}^{c}\frac{\sin t(x-a)-\sin t(x-b)}{t}\mathrm{d}t \\ &= \frac{1}{2\pi}\int_{-c(x-a)}^{c(x-a)}\frac{\sin v}{v}\mathrm{d}v - \frac{1}{2\pi}\int_{-c(x-b)}^{c(x-b)}\frac{\sin u}{u}\mathrm{d}u. \end{aligned}\tag{30}$$

函数

$$g(s,t)=\int_s^t \frac{\sin v}{v}\mathrm{d}v$$

关于 s 和 t 一致连续, 且当 $s\downarrow -\infty$ 和 $t\uparrow\infty$ 时

$$g(s,t)\to\pi. \tag{31}$$

因此, 存在常数 C, 使对于一切 c 和 x, 有 $|\Psi_c(x)|<C<\infty$. 此外, 由 (30) 和 (31) 式, 可见

$$\Psi_c(x)\to\Psi(x), c\to\infty,$$

其中

$$\Psi(x)=\begin{cases}0, & 若\ x<a\ 或\ x>b,\\ 1/2, & 若\ x=a\ 或\ x=b,\\ 1, & 若\ a<x<b.\end{cases}$$

设 μ 是 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的测度, 且 $\mu(a,b]=F(b)-F(a)$. 那么, 运用控制收敛定理和 §3 练习题 1, 当 $c\to\infty$ 时, 有

$$\begin{aligned}\Phi_c &= \int_{-\infty}^{\infty}\Psi_c(x)\mathrm{d}F(x)\to\int_{-\infty}^{\infty}\Psi(x)\mathrm{d}F(x)=\mu(a,b)+\frac{1}{2}\mu\{a\}+\frac{1}{2}\mu\{b\}\\ &= F(b-)-F(a)+\frac{1}{2}[F(a)-F(a-)+F(b)-F(b-)]\\ &= \frac{F(b)+F(b-)}{2}-\frac{F(a)+F(a-)}{2}=F(b)-F(a),\end{aligned}$$

其中最后一个等式对于函数 $F(x)$ 的一切连续点 a 和 b 成立.

于是, (25) 式得证.

b) 设 $\int_{-\infty}^{\infty}|\varphi(t)|\mathrm{d}t<\infty$. 记

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-itx}\varphi(t)\mathrm{d}t.$$

由控制收敛定理知, 此函数对 x 连续, 故在区间 $[a,b]$ 上可积. 所以仍可以应用傅比

尼定理, 可得对于一切函数 $F(x)$ 的一切连续点 a 和 b, 有

$$\begin{aligned}\int_a^b f(x)\mathrm{d}x &= \int_a^b \frac{1}{2\pi}\left[\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\varphi(t)\mathrm{d}t\right]\mathrm{d}x \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)\left[\int_a^b \mathrm{e}^{-itx}\mathrm{d}x\right]\mathrm{d}t = \lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\varphi(t)\left[\int_a^b \mathrm{e}^{-itx}\mathrm{d}x\right]\mathrm{d}t \\ &= \lim_{c\to\infty}\int_{-c}^{c}\frac{1}{2\pi}\frac{\mathrm{e}^{-ita}-\mathrm{e}^{-itb}}{it}\varphi(t)\mathrm{d}t = F(b)-F(a).\end{aligned}$$

由此可见,

$$F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y, \quad x \in \mathbb{R},$$

而且因为 $f(x)$ 连续, 且 $F(x)$ 是不减函数, 所以 $f(x)$ 是 $F(x)$ 的密度. □

注 逆转公式 (25) 给出了定理 2 的另一证明.

定理 4 随机向量 $\xi = (\xi_1, \cdots, \xi_n)$ 的分量独立的充分和必要条件是, 其特征函数等于各分量特征函数的乘积:

$$\mathbf{E}\mathrm{e}^{i(t_1\xi_1+\cdots+t_n\xi_n)} = \prod_{k=1}^{n}\mathbf{E}\mathrm{e}^{it_k\xi_k}, (t_1, \cdots, t_n) \in \mathbb{R}^n.$$

证明 由练习题 1, 可见必要性成立. 为证明充分性, 记 $F(x_1, \cdots, x_n)$ 为随机向量 $\xi = (\xi_1, \cdots, \xi_n)$ 的分布函数, 而 $F_k(x)$ 为 $\xi_k (1 \leqslant k \leqslant n)$ 的分布函数; $G = G(x_1, \cdots, x_n) = F_1(x_1)\cdots F_n(x_n)$. 那么, 根据傅比尼定理, 对于一切 $(t_1, \cdots, t_n) \in \mathbb{R}^n$, 有

$$\begin{aligned}&\int_{\mathbb{R}^n} \mathrm{e}^{i(t_1x_1+\cdots+t_nx_n)}\mathrm{d}G(x_1, \cdots, x_n) = \prod_{k=1}^{n}\int_{\mathbb{R}}\mathrm{e}^{it_kx_k}\mathrm{d}F_k(x_k) \\ &= \prod_{k=1}^{n}\mathbf{E}\mathrm{e}^{it_k\xi_k} = \mathbf{E}\mathrm{e}^{i(t_1\xi_1+\cdots+t_n\xi_n)} = \int_{\mathbb{R}^n}\mathrm{e}^{i(t_1x_1+\cdots+t_nx_n)}\mathrm{d}F(x_1, \cdots, x_n).\end{aligned}$$

于是, 由定理 2 (确切地说, 由定理 2 的多元类似; 亦见练习题 3) 可见, 根据 §5 的定理, 随机变量 $\xi_1, \cdots, \xi_n$ 独立. □

6. 特征函数的必要条件 在定理 1 中包含特征函数的某些必要条件. 例如, 假如函数 $\varphi = \varphi(t)$ 不满足该定理的前 3 个条件之一, 则这一函数就不是特征函数.

验证我们所感兴趣的函数 $\varphi = \varphi(t)$ 是否特性函数, 是一项复杂的事. 下面给出这方面一些结果 (不加证明).

定理 (博赫纳 – 辛钦 [S. Bochner–A. Я. Хинчин]) 设 $\varphi(t), t \in \mathbb{R}$, 是连续函数且 $\varphi(0) = 1$.$\varphi(t)$ 是特征函数的充分和必要条件是, 它是非负定函数, 即对于任意实数 $t_1, \cdots, t_n$ 和对于任意复数 $\lambda_1, \cdots, \lambda_n, n = 1, 2, \cdots$, 有

$$\sum_{i,j=1}^{n}\varphi(t_i - t_j)\lambda_i\bar{\lambda}_j \geqslant 0. \tag{32}$$

条件 (32) 的必要性显然: 因为如果

$$\varphi(t)=\int_{-\infty}^{\infty}\mathrm{e}^{itx}\mathrm{d}F(x),$$

则

$$\sum_{i,j=1}^{n}\varphi(t_i-t_j)\lambda_i\overline{\lambda}_j=\int_{-\infty}^{\infty}\left|\sum_{k=1}^{n}\lambda_k\mathrm{e}^{it_kx}\right|^2\mathrm{d}F(x)\geqslant 0.$$

条件 (32) 充分性的证明比较困难.(参见 [69] 的T.2, X IX.2).

定理 (波利亚 [G. Polia])　设 $\varphi(t)$ 是连续函数、偶函数、在 $(-\infty,0)$ 上 (因此在 $(0,\infty)$ 上) 是凹函数 (向下凸函数), 且满足条件: $\varphi(t)\geqslant 0,\varphi(0)=1,\varphi(t)\to 0(t\to\infty)$, 那么, $\varphi(t)$ 是特征函数 (参见 [69] 的 T.2, X V.2).

这一定理提供了构造特征函数的相当方便的方法. 例如, 函数

$$\varphi_1(t)=\mathrm{e}^{-|t|},$$

$$\varphi_2(t)=\begin{cases}1-|t|, & \text{若 } |t|\leqslant 1,\\ 0, & \text{若 } |t|>1.\end{cases}$$

就是这样的函数.

图 34 所表示的函数 $\varphi_3(x)$ 也是特征函数. 在区间 $[-a,a]$ 上 $\varphi_3(x)=\varphi_2(x)$. 然而, 与它们对应的分布函数 F_2 和 F_3 显然不同. 这个例子表明, 在有限区间上特征函数相同, 一般分布函数未必相同.

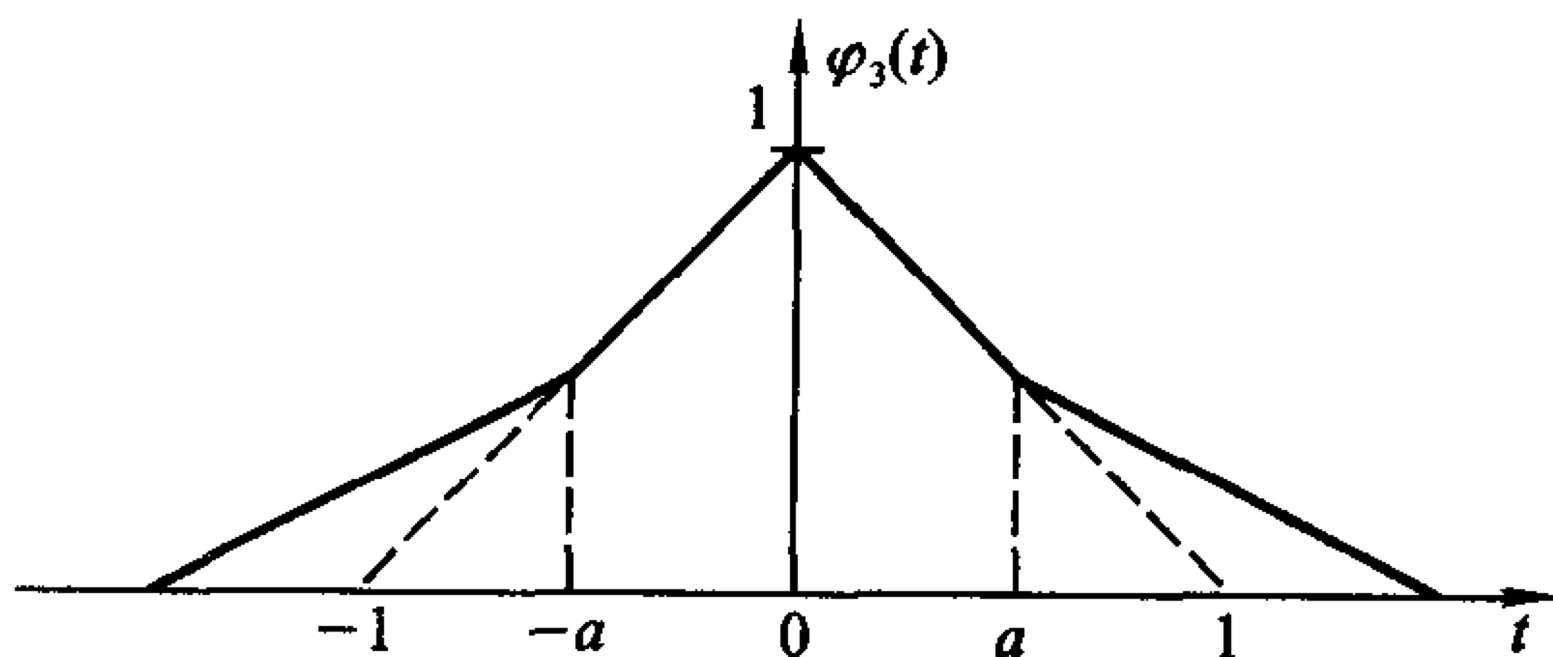


图 34

定理 (马钦凯维奇 [J. Marcinkiewicz])　如果特征函数 $\varphi(t)$ 具有形式 $\exp\mathscr{P}(t)$, 其中 $\mathscr{P}(t)$ 是多项式, 则此多项式的次数不可能大于 2(见 [135, 7.3]).

例如, 由此定理可见, 函数 $\exp\{-t^4\}$ 不是特征函数.

7. 某些特殊分布的特征函数　下面的定理是用例子表明, 如何根据随机变量的特征函数, 作出关于此变量的不平凡的论断.

定理 5　设 $\varphi_\xi(t)$ 是随机变量 ξ 的特征函数.

a) 如果对于某个 $t_0 \neq 0, |\varphi_\xi(t_0)| = 1$, 则 ξ 是步长为 $h = 2\pi/|t_0|$ 的格点随机变量, 即

$$\sum_{n=-\infty}^{\infty} \mathbf{P}\{\xi = a + nh\} = 1, \tag{33}$$

其中 a 是常数.

b) 如果对于两个不同的点 t 和 αt, 其中 α 是无理数, 有 $|\varphi_\xi(t)| = |\varphi_\xi(\alpha t)| = 1$, 则随机变量 ξ 是退化的:

$$\mathbf{P}\{\xi = a\} = 1,$$

其中 a 是常数.

c) 如果 $|\varphi_\xi(t)| \equiv 1$, 则随机变量 ξ 是退化的.

证明 a) 如果 $|\varphi_\xi(t_0)| = 1, t_0 \neq 0$, 则存在实数 a, 使 $t_0\varphi(t_0) = \mathrm{e}^{it_0a}$. 那么,

$$\mathrm{e}^{it_0a} = \int_{-\infty}^{\infty} \mathrm{e}^{-it_0x}\mathrm{d}F(x) \Rightarrow 1 = \int_{-\infty}^{\infty} \mathrm{e}^{-it_0(x-a)}\mathrm{d}F(x) \Rightarrow$$
$$\Rightarrow 1 = \int_{-\infty}^{\infty} \cos t_0(x-a)\mathrm{d}F(x) \Rightarrow \int_{-\infty}^{\infty} [1 - \cos t_0(x-a)]\mathrm{d}F(x) = 0.$$

由于 $1 - \cos t_0(x-a) \geqslant 0$, 由 §6 第 2 小节性质 **H**, 可见 ($\mathbf{P}$ – a.c.) 有

$$1 = \cos t_0(\xi - a),$$

而这与 (33) 式等价.

b) 由假设 $|\varphi_\xi(t)| = |\varphi_\xi(\alpha t)| = 1$ 和 (33) 式, 可见

$$\sum_{n=-\infty}^{\infty} \mathbf{P}\left\{\xi = a + \frac{2\pi n}{t}\right\} = \sum_{m=-\infty}^{\infty} \mathbf{P}\left\{\xi = b + \frac{2\pi m}{\alpha t}\right\} = 1.$$

如果 ξ 是非退化的, 则在集合

$$\left\{a + \frac{2\pi n}{t}, n = 0, \pm 1, \cdots\right\} \text{和} \left\{b + \frac{2\pi m}{\alpha t}, m = 0, \pm 1, \cdots\right\}$$

上至少存在两个不同的点:

$$a + \frac{2\pi n_1}{t} = b + \frac{2\pi m_1}{\alpha t}, \quad a + \frac{2\pi n_2}{t} = b + \frac{2\pi m_2}{\alpha t},$$

由此得

$$\frac{2\pi}{t}(n_1 - n_2) = \frac{2\pi}{\alpha t}(m_1 - m_2),$$

而这与假设 "α 是无理数" 矛盾, 故 b) 正确. 此外, 由命题 b) 可得命题 c). □

8. 随机变量的半不变量和矩的联系　设 $\xi = (\xi_1, \cdots, \xi_k)$ 是随机向量, 函数,

$$\varphi_\xi(\boldsymbol{t}) = \mathbf{E}e^{i(t,\xi)}, \quad \boldsymbol{t} = (t_1, \cdots, t_k),$$

是它的特征函数. 假设对于某个 $n \geqslant 1, \mathbf{E}|\xi_i|^n < \infty, i = 1, \cdots, k$. 由 §6 的赫尔德不等式 (29) 和李雅普诺夫不等式 (27), 可见对于一切非负 $\nu_1, \cdots, \nu_k$, 且 $\nu_1 + \cdots + \nu_k \leqslant n$,(混合) 矩 $\mathbf{E}(\xi_1^{\nu_1} \cdots \xi_k^{\nu_k})$ 存在.

如定理 1, 由此对于 $\nu_1 + \cdots + \nu_k \leqslant n$, 可见偏导数

$$\frac{\partial^{\nu_1 + \cdots + \nu_k}}{\partial t_1^{\nu_1} \cdots \partial t_k^{\nu_k}} \varphi_\xi(t_1, \cdots, t_k)$$

的存在性和连续性. 那么, 将 $\varphi_\xi(t_1, \cdots, t_k)$ 展为泰勒级数, 可得

$$\varphi_\xi(t_1, \cdots, t_k) = \sum_{\nu_1 + \cdots + \nu_k \leqslant n} \frac{i^{\nu_1 + \cdots + \nu_k}}{\nu_1! \cdots \nu_k!} m_\xi^{(\nu_1, \cdots, \nu_k)} t_1^{\nu_1} \cdots t_k^{\nu_k} + o(|t|^n), \tag{34}$$

其中 $|t| = |t_1| + \cdots + |t_k|$, 而

$$m_\xi^{(\nu_1, \cdots, \nu_k)} = \mathbf{E}\xi_1^{\nu_1} \cdots \xi_k^{\nu_k}$$

是 $\nu = (\nu_1, \cdots, \nu_k)$ 阶 (混合) 矩.

函数 $\varphi_\xi(t_1, \cdots, t_k)$ 连续, $\varphi_\xi(0, \cdots, 0) = 1$, 因此在 $0 = (0, \cdots, 0)$ 的某邻域内 $(|t| < \delta), \varphi_\xi(t_1, \cdots, t_k) \neq 0$. 在此邻域内存在连续偏导数

$$\frac{\partial^{\nu_1 + \cdots + \nu_k}}{\partial t_1^{\nu_1} \cdots \partial t_k^{\nu_k}} \ln \varphi_\xi(t_1, \cdots, t_k),$$

其中将 $\ln z$ 理解为对数的主值 (假如 $z = re^{i\theta}, -\pi < \theta \leqslant \pi$, 则假定 $\ln z$ 为 $\ln r + i\theta$).

因此 $\ln \varphi_\xi(t_1, \cdots, t_k)$ 可以表示为泰勒公式

$$\ln \varphi_\xi(t_1, \cdots, t_k) = \sum_{\nu_1 + \cdots + \nu_k \leqslant n} \frac{i^{\nu_1 + \cdots + \nu_k}}{\nu_1! \cdots \nu_k!} s_\xi^{(\nu_1, \cdots, \nu_k)} t_1^{\nu_1} \cdots t_k^{\nu_k} + o(|t|^n), \tag{35}$$

其中系数 $s_\xi^{(\nu_1, \cdots, \nu_k)}$ 称做随机向量 $\xi = (\xi_1, \cdots, \xi_k)$ 的 $\nu = (\nu_1, \cdots, \nu_k)$ 阶 (混合) *半不变量*或*累积量*.

注意, 如果 ξ 和 η 是两个独立随机向量, 则

$$\ln \varphi_{\xi+\eta}(t) = \ln \varphi_\xi(t) + \ln \varphi_\eta(t), \tag{36}$$

因此

$$s_{\xi+\eta}^{(\nu_1, \cdots, \nu_k)} = s_\xi^{(\nu_1, \cdots, \nu_k)} + s_\eta^{(\nu_1, \cdots, \nu_k)}. \tag{37}$$

(也正是这一条性质说明 "*半不变量*" 名称的合理性.)

为简化记号, 并且使 (34) 和 (35) 式具有 "一维" 的形式, 我们引进如下记号. 如果 $\nu=(\nu_1,\cdots,\nu_k)$ 是具有非负整数分量的向量, 则设

$$\nu!=\nu_1!\cdots\nu_k!,\quad |\nu|=\nu_1+\cdots+\nu_k,\quad t^\nu=t_1^{\nu_1}\cdots t_k^{\nu_k}.$$

亦设

$$s_\xi^{(\nu)}=s_\xi^{(\nu_1,\cdots,\nu_k)},m_\xi^{(\nu)}=m_\xi^{(\nu_1,\cdots,\nu_k)}.$$

那么,(34) 和 (35) 式具有如下形式:

$$\varphi_\xi(t)=\sum_{|\nu|\leqslant n}\frac{i^{|\nu|}}{\nu!}m_\xi^{(\nu)}t^\nu+o(|t|^n). \tag{38}$$

$$\ln\varphi_\xi(t)=\sum_{|\nu|\leqslant n}\frac{i^{|\nu|}}{\nu!}s_\xi^{(\nu)}t^\nu+o(|t|^n). \tag{39}$$

下面的定理及其系给出矩和半不变量的联系公式.

定理 6 设 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_k)$ 是随机向量, 满足 $\mathbf{E}|\xi_i|^n<\infty,i=1,\cdots,k,n\geqslant 1$. 那么, 对于一切 $\nu=(\nu_1,\cdots,\nu_k)$ 且 $|\nu|\leqslant n$, 有

$$m_\xi^{(\nu)}=\sum_{\lambda^{(1)}+\cdots+\lambda^{(q)}=\nu}\frac{1}{q!}\frac{\nu!}{\lambda^{(1)}!\cdots\lambda^{(q)}!}\prod_{p=1}^q s_\xi^{(\lambda^{(p)})}, \tag{40}$$

$$s_\xi^{(\nu)}=\sum_{\lambda^{(1)}+\cdots+\lambda^{(q)}=\nu}\frac{(-1)^{q-1}}{q!}\frac{\nu!}{\lambda^{(1)}!\cdots\lambda^{(q)}!}\prod_{p=1}^q m_\xi^{(\lambda^{(p)})}, \tag{41}$$

其中 $\sum\limits_{\lambda^{(1)}+\cdots+\lambda^{(q)}=\nu}$ 表示对于一切满足 $|\lambda^{(p)}|>0,\lambda^{(1)}+\cdots+\lambda^{(q)}=\nu$ 的 "有序非负整数分量的向量 $\lambda^{(p)}$" 求和.

证明 由于

$$\varphi_\xi(t)=\exp\{\ln\varphi_\xi(t)\},$$

则将 $\exp\{\cdots\}$ 展成泰勒级数并考虑到 (39) 式, 得

$$\varphi_\xi(t)=1+\sum_{q=1}^n\frac{1}{q!}\left(\sum_{1\leqslant|\lambda|\leqslant n}\frac{i^{|\lambda|}}{\lambda!}s_\xi^{(\lambda)}t^\lambda\right)^q+o(|t|^n). \tag{42}$$

比较 (38) 和 (42) 式右侧 t^λ 的项, 并考虑到 $|\lambda^{(1)}|+\cdots+|\lambda^{(q)}|=|\lambda^{(1)}+\cdots+\lambda^{(q)}|$, 得 (40) 式.

其次,

$$\ln\varphi_\xi(t)=\ln\left[1+\sum_{1\leqslant|\lambda|\leqslant n}\frac{i^{|\lambda|}}{\lambda!}m_\xi^{(\lambda)}t^\lambda+o(|t|^n)\right]. \tag{43}$$

对较小的 z 有如下展开式

$$\ln(1+z)=\sum_{q=1}^{n}\frac{(-1)^{q-1}}{q}z^q+o(z^n).$$

将此展开式用于 (43) 式, 然后将项t^λ的系数与 (38) 式右侧相应项的系数比较, 得 (41) 式. □

系 1 *在矩和半不变量之间有下列公式:*

$$m_\xi^{(\nu)}=\sum_{\{r_1\lambda^{(1)}+\cdots+r_x\lambda^{(x)}=\nu\}}\frac{1}{r_1!\cdots r_x!}\frac{\nu!}{(\lambda^{(1)}!)^{r_1}\cdots(\lambda^{(x)}!)^{r_x}}\prod_{j=1}^{x}[s_\xi^{(\lambda^{(j)})}]^{r_j},\tag{44}$$

$$s_\xi^{(\nu)}=\sum_{\{r_1\lambda^{(1)}+\cdots+r_x\lambda^{(x)}=\nu\}}\frac{(-1)^{q-1}(q-1)!}{r_1!\cdots r_x!}\frac{\nu!}{(\lambda^{(1)}!)^{r_1}\cdots(\lambda^{(x)}!)^{r_x}}\prod_{j=1}^{x}[m_\xi^{(\lambda^{(j)})}]^{r_j},\tag{45}$$

其中 $\sum_{\{r_1\lambda^{(1)}+\cdots+r_x\lambda^{(x)}=\nu\}}$ *表示对于一切满足* $|\lambda^{(j)}|>0, r_1\lambda^{(1)}+\cdots+r_x\lambda^{(x)}=\nu$ *的"有序正整数* r_j *的数组"求和.*

为证明 (44) 式, 假设 (40) 式中相应地有 r_1 个向量等于 $\lambda^{(i_1)},\cdots,r_x$ 个向量等于 $\lambda^{(i_x)}(r_j>0,r_1+\cdots+r_x=q)$, 而且所有向量 $\lambda^{(i_s)}$ 两两不同. 恰好存在 $q!/(r_1!\cdots r_x!)$ 个不同的向量组 (精确到顺序) 与向量组 $(\lambda^{(1)},\cdots,\lambda^{(q)})$ 重合. 然而, 假如两个向量组, 例如, $(\lambda^{(1)},\cdots,\lambda^{(q)})$ 和 $(\overline{\lambda}^{(1)},\cdots,\overline{\lambda}^{(q)})$ 仅区别于分量的顺序, 则

$$\prod_{p=1}^{q}s_\xi^{(\lambda^{(p)})}=\prod_{p=1}^{q}s_\xi^{(\overline{\lambda}^{(p)})}.$$

所以, 可以将仅区别于顺序的数组视为同一数组, 则由 (40) 式得 (44) 式.

类似地可以由 (41) 式得 (45) 式.

系 2 考虑 $\nu=(1,\cdots,1)$ 的特殊情形. 这时, 矩 $m_\xi^{(\nu)}\equiv\mathbf{E}\xi_1\cdots\xi_k$ 以及相应的半不变量称做**简单的**.

联系简单矩和半不变量的公式, 可以由上面给出的公式得到. 不过, 最好将其写为另一种形式.

为此引进下面的记号.

设 $I_\xi=\{1,2,\cdots,k\}$ 是向量 $\xi=(\xi_1,\cdots,\xi_k)$ 之分量下标的集合. 假如 $I\subseteq I_\xi$, 则以 ξ_I 表示由下标属于 I 的向量 $\xi=(\xi_1,\cdots,\xi_k)$ 之分量构成的向量. 设 $\chi(I)$ 表示向量 $\{\chi_1,\cdots,\chi_n\}$, 其中, 若 $i\in I$, 则 $\chi_i=1$; 若 $i\notin I$, 则 $\chi_i=0$. 这些向量与集合 $I\subseteq I_\xi$ 一一对应. 因此, 记

$$m_\xi(I)=m_\xi^{(\chi(I))},\quad s_\xi(I)=s_\xi^{(\chi(I))}.$$

换句话说, $m_\xi(I)$ 和 $s_\xi(I)$ 是向量 $\xi=(\xi_1,\cdots,\xi_k)$ 之子向量, ξ_I 是简单矩和半不变量.

其次, 根据定义 (见第一章 §1 第 3 小节), 集合 I 的分割, 指不相交非空集合 I_p 组, 如果 $\sum_p I_p = I$.

鉴于这些记号, 有如下公式:

$$m_\xi(I) = \sum_{\sum_{p=1}^q I_p = I} \prod_{p=1}^q s_\xi(I_p), \tag{46}$$

$$s_\xi(I) = \sum_{\sum_{p=1}^q I_p = I} (-1)^{q-1}(q-1)! \prod_{p=1}^q m_\xi(I_p), \tag{47}$$

其中 $\sum\limits_{\sum_{p=1}^q I_p = I}$ 表示对于集合 I 的一切分割求和, 其中 $1 \leqslant q \leqslant N(I)$, 而 $N(I)$ 是集合 I 中的元素个数.

为证明 (46) 式, 要借助 (44) 式. 如果 $\nu = \chi(I)$ 和 $\lambda^{(1)} + \cdots + \lambda^{(q)} = \nu$, 则 $\lambda^{(p)} = \chi(I_p), I_p \subseteq I$, 一切 $\lambda^{(p)}$ 各不相同, $\lambda^{(p)}! = \nu! = 1$, 且每一无序组 $\{\chi(I_1), \cdots, \chi(I_q)\}$ 与分割 $I = \sum_{p=1}^q I_p$ 一一对应. 从而, 由 (44) 式得 (46) 式.

类似地, 可以由 (45) 式证明表达式 (47) 的正确性.

例 4 设 ξ 是随机变量 $(k=1), m_n = m_\xi^{(n)} = \mathbf{E}\xi^n, s_n = s_\xi^{(n)}$. 那么, 由 (40) 和 (41) 式得如下公式:

$$\begin{aligned} &m_1 = s_1, \\ &m_2 = s_2 + s_1^2, \\ &m_3 = s_3 + 3s_1s_2 + s_1^3, \\ &m_4 = s_4 + 3s_2^2 + 4s_1s_3 + 6s_1^2s_2 + s_1^4, \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{48}$$

和

$$\begin{aligned} &s_1 = m_1 = \mathbf{E}\xi, \\ &s_2 = m_2 - m_1^2 = \mathbf{D}\xi, \\ &s_3 = m_3 - 3m_1m_2 + 2m_1, \\ &s_4 = m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4, \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{49}$$

例 5 设 $\xi \sim N(m, \sigma^2)$, 由 (9) 式可见

$$\ln \varphi_\xi(t) = itm - \frac{t^2\sigma^2}{2}.$$

由 (39) 式 $s_1 = m, s_2 = \sigma^2$, 而从 3 阶起一切半不变量都等于 0, 即 $s_n = 0 (n \geqslant 3)$.

需要指出, 由于马钦凯维奇定理, 形如 $\exp \mathscr{P}(t)$ 的函数, 其中 $\mathscr{P}(t)$ 是多项式, 则只有当该多项式的次数不大于 2 时, 才可以做特征函数. 特别, 由此可见高斯分布是具有下面性质的唯一分布: 从某个 $n \geqslant 3$ 起一切半不变量 $s_n = 0$.

例 6 设 ξ 是参数为 $\lambda>0$ 的泊松随机变量, 则根据 (11) 式,

$$\ln \varphi_\xi(t)=\lambda(\mathrm{e}^{it}-1).$$

由此可见, 对于一切 $n\geqslant 1$, 有

$$s_n=\lambda. \tag{50}$$

例 7 设 $\xi=(\xi_1,\cdots,\xi_k)$ 是随机向量, 则

$$\begin{aligned}
m_\xi(1)&=s_\xi(1),\\
m_\xi(1,2)&=s_\xi(1,2)+s_\xi(1)s_\xi(2),\\
m_\xi(1,2,3)&=s_\xi(1,2,3)+s_\xi(1,2)s_\xi(3)+s_\xi(1,3)s_\xi(2),\\
&\quad+s_\xi(2,3)s_\xi(1)+s_\xi(1)s_\xi(2)s_\xi(3),\\
&\cdots\cdots\cdots\cdots
\end{aligned} \tag{51}$$

这些公式表明, 简单矩可以相当对称地通过简单半不变量表示. 假如设 $\xi_1\equiv\xi_2\equiv\cdots\equiv\xi_k$, 则由 (51) 式自然地得到 (48) 式.

由 (51) 式可见公式 (48) 中系数的 "组群" 来源. 此外, 由 (51) 式可见,

$$s_\xi(1,2)=m_\xi(1,2)-m_\xi(1)m_\xi(2)=\mathbf{E}\xi_1\xi_2-\mathbf{E}\xi_1\mathbf{E}\xi_2, \tag{52}$$

即 $s_\xi(1,2)$ 恰好是随机变量 ξ_1 和 ξ_2 的协方差.

9. 矩问题的唯一性 设随机变量 ξ 的分布函数为 $F(x)$, 特征函数为 $\varphi(t)$. 假设有一切阶矩 $m_n=\mathbf{E}\xi^n, n\geqslant 1$.

由定理 2 知特征函数唯一决定概率分布. 现在提出另一个问题 (矩的唯一性问题): 矩 $\{m_n\}_{n\geqslant 1}$ 是否唯一决定概率分布?

更确切地说, 假设两个分布函数 F 和 G 所有的矩都相同, 即对于一切整数 $n\geqslant 0$, 有

$$\int_{-\infty}^{\infty}x^n\mathrm{d}F(x)=\int_{-\infty}^{\infty}x^n\mathrm{d}G(x). \tag{53}$$

问由此是否可以得出 F 和 G 相同的结论?

一般, 对该问题的回答是否定的. 为证实这一点, 我们考虑分布函数 F, 假设其密度为

$$f(x)=\begin{cases}k\mathrm{e}^{-\alpha x^\lambda}, & 若\ x>0,\\ 0, & 若\ x\leqslant 0,\end{cases}$$

其中 $\alpha>0, 0<\lambda<1/2$, 而常数 k 由规范性条件 $\int_0^\infty f(x)\mathrm{d}x=1$ 决定.

记 $\beta=\alpha\tan\lambda\pi$, 则

$$g(x)=\begin{cases}k\mathrm{e}^{-\alpha x^\lambda}[1+\varepsilon\sin(\beta x^\lambda)], & |\varepsilon|<1,\quad x>0.\\ 0, & 若\ x\leqslant 0\end{cases}$$

显然 $g(x) \geqslant 0$. 现在验证对于所有整数 $n \geqslant 0$,

$$\int_0^{\infty} x^n \mathrm{e}^{-\alpha x^{\lambda}} \sin \beta x^{\lambda} \mathrm{d}x = 0. \tag{54}$$

熟知, 对于 $p > 0$ 和复数 q 的实部 $\mathrm{Re}\, q > 0$, 有

$$\int_0^{\infty} t^{p-1} \mathrm{e}^{-qt} \mathrm{d}t = \frac{\Gamma(p)}{q^p}.$$

设 $p = (n+1)/\lambda, q = \alpha + i\beta, t = x^{\lambda}$, 则

$$\begin{aligned}
&\int_0^{\infty} x^{\lambda\left(\frac{n+1}{\lambda}-1\right)} \mathrm{e}^{-(\alpha+i\beta)x^{\lambda}} \lambda x^{\lambda-1} \mathrm{d}x = \lambda \int_0^{\infty} x^n \mathrm{e}^{-(\alpha+i\beta)x^{\lambda}} \mathrm{d}x \\
&= \lambda \int_0^{\infty} x^n \mathrm{e}^{-\alpha x^{\lambda}} \cos \beta x^{\alpha} \mathrm{d}x - i\lambda \int_0^{\infty} x^n \mathrm{e}^{-\alpha x^{\lambda}} \sin \beta x^{\lambda} \mathrm{d}x \\
&= \frac{\Gamma\left(\frac{n+1}{\lambda}\right)}{\alpha^{\frac{n+1}{\lambda}} (1 + i \tan \lambda\pi)^{\frac{n+1}{\lambda}}}.
\end{aligned} \tag{55}$$

但是, 由于 $\sin(n+1)\pi = 0$, 有

$$\begin{aligned}
(1 + i \tan \lambda\pi)^{\frac{n+1}{\lambda}} &= (\cos \lambda\pi + i \sin \lambda\pi)^{\frac{n+1}{\lambda}} (\cos \lambda\pi)^{-\frac{n+1}{\lambda}} \\
&= \mathrm{e}^{i(n+1)\pi} (\cos \lambda\pi)^{-\frac{n+1}{\lambda}} = \cos(n+1)\pi \times (\cos \lambda\pi)^{-\frac{n+1}{\lambda}}.
\end{aligned}$$

这样,(55) 式的右侧是实的, 即对于任意整数 $n \geqslant 0$, (54) 式成立. 现在 $G(x)$, 作为取以 $g(x)$ 为密度的分布函数. 那么, 由 (54) 式知分布函数 F 和 G 的一切矩对应相等, 即对于任意整数 $n \geqslant 0$, (53) 式成立.

现在引进保障矩问题唯一性的某些充分条件.

定理 7 假设 $F = F(x)$ 是分布函数, 且对于 $n \geqslant 1$,

$$\mu_n = \int_{-\infty}^{\infty} |x|^n \mathrm{d}F(x).$$

如果

$$\varlimsup_{n\to\infty} \frac{\mu_n^{1/n}}{n} < \infty, \tag{56}$$

则矩 $\{m_n\}_{n\geqslant 1}$, 其中

$$m_n = \int_{-\infty}^{\infty} x^n \mathrm{d}F(x),$$

唯一地决定分布函数 $F = F(x)$.

证明 由 (56) 式和定理 1 的命题 7), 可见存在 $t_0 > 0$, 使对于一切 $|t| \leqslant t_0$, 特征函数 $\varphi(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} \mathrm{d}F(x)$ 可以表示为

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} m_k,$$

因而矩 $\{m_n\}_{n\geqslant 1}$ 对于一切 $|t|\leqslant t_0$, 唯一决定特征函数 $\varphi(t)$ 的值.

取点 $s,|s|\leqslant t_0/2$. 那么, 如同 (15) 式的证明, 由 (56) 式可以导出, 对于一切 $|t-s|\leqslant t_0$,

$$\varphi(t)=\sum_{k=0}^{\infty}\frac{(t-s)^k}{k!}\varphi^{(k)}(s),$$

其中

$$\varphi^{(k)}(s)=i^k\int_{-\infty}^{\infty}x^k\mathrm{e}^{isx}\mathrm{d}F(x)$$

唯一决定于矩 $\{m_n\}_{n\geqslant 1}$. 从而, 对于一切 $|t|\leqslant 3t_0/2$, 这些矩唯一决定特征函数 $\varphi(t)$ 的值. 继续这一过程可以断定, 对于一切 t, 矩 $\{m_n\}_{n\geqslant 1}$ 唯一决定特征函数 $\varphi(t)$, 故也唯一决定分布函数 $F(x)$. □

系 1 矩唯一决定集中在有限区间上的概率分布.

系 2 矩问题唯一性的充分条件是:

$$\overline{\lim_{n\to\infty}}\frac{(m_{2n})^{1/(2n)}}{2n}<\infty. \tag{57}$$

为证明只需注意到, 对于奇数阶矩分部估计, 然后利用条件 (56).

例 设 $F(x)$ 是正态分布函数,

$$F(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2\sigma^2}}\mathrm{d}t.$$

那么,

$$m_{2n+1}=0,\quad m_{2n}=\frac{(2n)!}{2^n n!}\sigma^{2n},$$

而由 (57) 式, 可见这些矩是只有正态分布才有的矩.

最后, 我们引进卡尔莱曼 (T.Carleman) 准则 (不加证明).

卡尔莱曼准则 (矩问题的唯一性) [69, 卷 2, Ⅶ.3.]

a) 设 $\{m_n\}_{n\geqslant 1}$ 是某概率分布的矩, 且

$$\sum_{n=0}^{\infty}\frac{1}{(m_{2n})^{1/(2n)}}=\infty.$$

那么, $\{m_n\}_{n\geqslant 1}$ 唯一决定概率分布.

b) 如果 $\{m_n\}_{n\geqslant 1}$ 是集中在 $[0,\infty)$ 上概率分布的矩, 则 $\{m_n\}_{n\geqslant 1}$ 唯一性的充分条件是

$$\sum_{n=0}^{\infty}\frac{1}{(m_n)^{1/(2n)}}=\infty.$$

10. 埃森不等式 $F = F(x)$ 和 $G = G(x)$ 分别是以 $f = f(t)$ 和 $g = g(t)$ 为特征函数的分布函数. 下面的定理说明 (略去证明), 根据 f 和 g 的接近程度, 如何 (在均匀尺度下) 估计 F 和 G 的接近程度. (关于该定理的应用见第三章 §11.)

定理 (埃森 [C. G. Esseen]不等式) 设 $G(x)$ 有导数且 $\sup_x |G'(x)| \leqslant C$. 则

$$\sup_x |F(x) - G(x)| \leqslant \frac{2}{\pi} \int_0^T \left| \frac{f(t) - g(t)}{t} \right| \mathrm{d}t + \frac{24}{\pi T} \sup_x |G'(x)|.$$

11. 常见分布的特征函数表 表 2–4 和表 2–5 是一些常见分布的特征函数表:

表 2 – 4

分布名称	特征函数
离散均匀	$\frac{1}{N} \frac{\mathrm{e}^{it}}{1 - \mathrm{e}^{it}} (1 - \mathrm{e}^{itN})$
伯努利 (J. Bernoulli)	$q + p\mathrm{e}^{it}$
二项	$(q + p\mathrm{e}^{it})^n$
泊松 (S. D. Poisson)	$\exp\{\lambda(\mathrm{e}^{it} - 1)\}$
几何	$p(1 - q\mathrm{e}^{it})^{-1}$
负二项 (帕斯卡)	$[p(1 - q\mathrm{e}^{it})^{-1}]^r$

表 2 – 5

分布名称	特征函数
$[a, b]$ 上的均匀	$\frac{\mathrm{e}^{itb} - \mathrm{e}^{ita}}{it(b - a)}$
正态或高斯 (C. F. Gauss)	$\exp\left\{itm - \frac{\sigma^2 t^2}{2}\right\}$
伽马 (Γ)	$(1 - it\beta)^{-\alpha}$
贝塔 (B)	$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \sum_{k=0}^{\infty} \frac{(it)^k \Gamma(\alpha + k)}{k!\Gamma(\alpha + \beta + k)\Gamma(1 + k)}$
指数	$\frac{\lambda}{\lambda - it}$
双测指数	$\frac{\lambda^2 \mathrm{e}^{it\alpha}}{\lambda^2 + t^2}$
χ^2 (卡方)	$(1 - 2it)^{-n/2}$
t (学生)	$\frac{\sqrt{\pi}\Gamma((n + 1)/2)}{\Gamma(n/2)} \frac{\|t\| \exp\{-\sqrt{n}\}}{2^{2(m-1)}(m - 1)!} \sum_{k=0}^{m-1} (2k)!\mathrm{C}_{n-1+k}^{2k} (2\|t\|\sqrt{n})^{m-1-k}$, 若 $m = (n + 1)/2$ 是整数
柯西 (A. L. Cauchy)	$\mathrm{e}^{-\theta\|t\|}$

12. 练习题

1. 设 ξ 和 η 是独立随机变量, $f(x)=f_1(x)+if_2(x), g(x)=g_1(x)+ig_2(x)$, 其中 $f_k(x)$ 和 $g_k(x)(k=1,2)$ 是博雷尔函数. 证明, 如果 $\mathbf{E}|f(\xi)|<\infty, \mathbf{E}|g(\xi)|<\infty$, 则

$$\mathbf{E}|f(\xi)g(\eta)|<\infty$$

且

$$\mathbf{E}f(\xi)g(\eta)=\mathbf{E}f(\xi)\times\mathbf{E}g(\eta).$$

2. 设 $\xi=(\xi_1,\cdots,\xi_n)$ 且 $\mathbf{E}\|\xi\|^n<\infty$, 其中 $\|\xi\|=\sqrt{\sum\xi_i^2}$. 证明

$$\varphi_\xi(\boldsymbol{t})=\sum_{k=0}^{n}\frac{i^k}{k!}\mathbf{E}(\boldsymbol{t},\xi)^k+\varepsilon_n(\boldsymbol{t})\|\boldsymbol{t}\|^n,$$

其中 $\boldsymbol{t}=(t_1,\cdots,t_n)$ 且 $\varepsilon_n(\boldsymbol{t})\to 0, \boldsymbol{t}\to 0$.

3. 对于 n 维分布函数 $F=F_n(x_1,\cdots,x_n)$ 和 $G=G_n(x_1,\cdots,x_n)$, 证明定理 2.

4. 设 $F=F(x_1,\cdots,x_n)$ 是 n 维分布函数, 而 $\varphi=\varphi(t_1,\cdots,t_n)$ 是其特征函数. 利用 §3(12) 式的记号, 证明逆转公式:

$$\mathbf{P}(a,b]=\lim_{c\to\infty}\frac{1}{(2\pi)^n}\int_{-c}^{c}\cdots\int_{-c}^{c}\prod_{k=1}^{n}\frac{\mathrm{e}^{-it_ka_k}-\mathrm{e}^{-it_kb_k}}{it_k}\varphi(t_1,\cdots,t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n.$$

(在上面引进的逆转公式中, $(a,b]$ 是函数 $\mathbf{P}(a,b]$ 的连续区间, 即对于一切 $k=1,\cdots,n$, 点 a_k,b_k 是边缘分布函数 $F_k(x_k)$ 的连续点, 而边缘分布函数 $F_k(x_k)$ 是由联合分布函数 $F=F(x_1,\cdots,x_n)$ 将除 x_k 之外的其余变量设为 $+\infty$ 得到的.)

5. 设 $\varphi_k(t), k\geqslant 1$ 是特征函数, 而非负实数 $\lambda_k, k\geqslant 1$, 满足条件 $\sum\lambda_k=1$. 证明 $\sum\lambda_k\varphi_k(t)$ 是特征函数.

6. 设 $\varphi(t)$ 是特征函数, 问 $\mathrm{Re}\varphi(t)$ 和 $\mathrm{Im}\varphi(t)$ 是否特征函数?

7. 设 $\varphi_1,\varphi_2,\varphi_3$ 是特征函数, 且 $\varphi_1\varphi_2=\varphi_1\varphi_3$, 问是否由此可见 $\varphi_2=\varphi_3$?

8. 证明表 2 – 4 和表 2 – 5 中特征函数式的正确性.

9. 设 ξ 是整数值随机变量, 而 $\varphi(t)$ 是其特征函数. 证明

$$\mathbf{P}\{\xi=k\}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{e}^{-ikt}\varphi_\xi(t)\mathrm{d}t,\quad k=0,\pm1,\pm2,\cdots.$$

10. 证明在具有勒贝格测度 μ 的空间 $L^2=L^2((-\pi,\pi],\mathscr{B}(-\pi,\pi])$ 中, 函数系

$$\left\{\frac{1}{\sqrt{2\pi}}\mathrm{e}^{i\lambda n}, n=0,\pm1,\cdots\right\}$$

构成规范正交基.

11. 假设在博赫纳 – 辛钦定理中所考虑的函数 $\varphi(t)$ 是连续的. 证明下面 (里斯 [F. Riesz]) 结果, 该结果说明在什么程度上可以去掉连续性假设.

假设 $\varphi=\varphi(t)$ 是复数值勒贝格可测函数且 $\varphi(0)=1$. 那么, 函数 $\varphi(t)$ 是正定的, 当且仅当它 (在数轴上勒贝格测度几乎处处) 等于某一特征函数.

12. 判断下列各函数是否特征函数:

$$\varphi(t)=\mathrm{e}^{-|t|^k}, 0\leqslant k\leqslant 2;\quad \varphi(t)=\mathrm{e}^{-|t|^k}, k>2;$$

$$\varphi(t)=(1+|t|)^{-1};\quad \varphi(t)=(1+t^4)^{-1};$$

$$\varphi(t)=\begin{cases}1-|t|^3, & 若\ |t|\leqslant 1,\\ 0, & 若\ |t|>1;\end{cases}\qquad \varphi(t)=\begin{cases}1-|t|, & 若\ |t|\leqslant 1/2,\\ 1/(4|t|), & 若\ |t|>1/2.\end{cases}$$

13. 设 $\varphi(t)$ 是分布函数 $F=F(x)$ 的特征函数, $\{x_n\}$ 是函数 $F(x)[\Delta F(x_n)\equiv F(x_n)-F(x_n-)>0]$ 的间断点的集合. 证明

$$\lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T}|\varphi(t)|^2\mathrm{d}t=\sum_{n\geqslant 1}[\Delta F(x_n)]^2.$$

14. 称函数

$$Q(X;l)=\sup_{x\in\mathbb{R}}\mathbf{P}\{x\leqslant X\leqslant x+l\}$$

为随机变量 X 的集中函数. 证明:

(a) 若 X 和 Y 是独立随机变量, 则对于一切 $l\geqslant 0$,

$$Q(X+Y;l)\leqslant\min\{Q(X;l),Q(Y;l)\};$$

(b) 存在 x_l^*, 使随机变量 X 的 $Q(X;l)=\mathbf{P}\{x_l^*\leqslant X\leqslant x_l^*+l\}$ 和随机变量 X 的分布函数连续, 当且仅当 $Q(X;0)=0$.

15. 设 X 是分布函数为 $F=F(x)$ 的随机变量, $\{m_n\}_{n\geqslant 1}$ 是其矩序列, 其中 $m_k=\int_{-\infty}^{\infty}x^k\mathrm{d}F(x)$. 证明, 若对于某个 $s>0$, 级数 $\sum\limits_{k=1}^{\infty}\frac{m_k}{k!}s^k$ 绝对收敛, 则 $\{m_n\}_{n\geqslant 1}$ 唯一决定分布函数 $F=F(x)$.

16. 设 $F=F(x)$ 分布函数, 而 $\varphi(t)=\int_{-\infty}^{\infty}\mathrm{e}^{itx}\mathrm{d}F(x)$ 是其的特征函数. 证明

$$\lim_{c\to\infty}\frac{1}{2c}\int_{-c}^{c}\mathrm{e}^{-itx}\varphi(t)\mathrm{d}t=F(x)-F(x-),$$

$$\lim_{c\to\infty}\frac{1}{2c}\int_{-c}^{c}|\varphi(t)|^2\mathrm{d}t=\sum_{x\in\mathbb{R}}[F(x)-F(x-)]^2.$$

17. 证明每一个特征函数 $\varphi(t)$ 都满足不等式: $1-\mathrm{Re}\varphi(2t)\leqslant 4[1-\mathrm{Re}\varphi(t)]$.

18. 假设特征函数 $\varphi(t)$ 满足: $\varphi(t)=1+f(t)+o(t^2), t\to 0$, 其中 $f(t)=-f(-t)$. 证明 $\varphi(t)\equiv 1$.

19. 证明对于每一个 $n \geqslant 1$, 函数

$$\varphi_n(t) = \frac{e^{it} - \sum_{k=0}^{n-1}(it)^k/k!}{(it)^n/n!}$$

是特征函数.

20. 证明

$$\frac{2}{\pi}\int_{-\infty}^{\infty}\frac{1-\mathrm{Re}\varphi(x)}{t^2}\mathrm{d}t = \int_{-\infty}^{\infty}|x|\mathrm{d}F(x).$$

21. 假设特征函数 $\varphi(t)$ 满足: $\varphi(t) = 1 + O(|t|^\alpha), t \to 0$, 其中 $\alpha \in (0,2]$. 证明以 $\varphi(t)$ 为特征函数的随机变量 ξ 具有如下性质:

$$\mathbf{P}\{|\xi| > x\} = O(x^{-\alpha}), \quad x \to 0.$$

22. 如果 $\varphi(t)$ 是特征函数, 则函数 $|\varphi(t)|^2$ 也是特征函数.

23. 设 X 和 Y 是独立同分布随机变量, 其数学期望为 0 和方差为 1. 基于特征函数的性质, 证明, 如果随机变量 $(X+Y)/\sqrt{2}$ 的分布等同于随机变量 X 和 Y 的分布 F, 则 F 是正态分布函数.

24. 如果 $\varphi = \varphi(t)$ 是特征函数, 则对于某一个 $\lambda \geqslant 0$ 函数 $e^{\lambda(\varphi-1)}$ 也是特征函数.

25. 设 X 是分布函数为 F 的非负随机变量. 对于任意 $\lambda \geqslant 0$, 由公式

$$\widehat{F}(\lambda) = \mathbf{E}e^{-\lambda X} = \int_{[0,\infty)} e^{-\lambda x}\mathrm{d}F(x)$$

定义的函数 $\widehat{F} = \widehat{F}(\lambda)$ 称做 X 的*拉普拉斯变换*. 证明如下伯恩斯坦 (С. Н. Бернштейн) 准则: 函数 $f = f(\lambda), \lambda \in (0,\infty)$ 是分布函数 $F = F(x), x \in [0,\infty)$ 的拉普拉斯变换的充分和必要条件是: 函数 $f = f(\lambda)$ 是完全单调的 (即存在所有阶导数 $f^{(n)}(\lambda), n \geqslant 0$ 且 $(-1)^n f^{(n)}(\lambda) \geqslant 0$).

26. 设 $\varphi(t)$ 是特征函数, 证明下面的函数也是特征函数:

$$\int_0^1 \varphi(ut)\mathrm{d}u, \quad \int_0^\infty e^{-u}\varphi(ut)\mathrm{d}u.$$

§13. 高斯系

1. 高斯系的特点和重要性 在概率论和数理统计中, 高斯分布或正态分布、高斯随机变量、高斯过程和高斯系有特别重要的作用. 这首先因为有*中心极限定理* (第三章 §4), 棣莫弗 – 拉普拉斯定理是其特殊情形 (第一章 §6). 根据这一定理, 正态分布具有通用的特点: 在并不 "拘谨" 的条件下, 大量独立随机变量或随机向量之和的分布, 可以很好地用正态分布来逼近.

正是这种情况, 从理论上说明了统计实践中普遍的 "误差律", 它表现为由大量独立的 "基本" 误差叠加形成的测量误差服从正态分布.

多维正态分布依赖于少量参数, 在建立简单的概率模型时是其毋容置疑的优点. 高斯随机变量有有限二阶矩, 故可以用希尔伯特空间方法研究其性质. 重要的情况是, 对于高斯随机变量的情形, 不相关性变为独立性, 使得有可能大为加强 "L^2- 理论" 的结果.

2. 高斯系的定义 注意到 (根据 §8) 随机变量 $\xi=\xi(\omega)$ 称做服从参数为 m 和 σ^2 的高斯分布或正态分布的 $(\xi\sim N(m,\sigma^2),|m|<\infty,\sigma^2>0)$, 如果其密度为

$$f_\xi(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}},\tag{1}$$

其中 $\sigma=+\sqrt{\sigma^2}$.

当 $\sigma\downarrow 0$ 时, 密度 $f_\xi(x)$ "收敛于集中在点 $x=m$ 的 δ- 函数". 所以自然称随机变量 $\xi=\xi(\omega)$ 服从参数为 m 和 $\sigma^2=0$ 的高斯分布或正态分布 $(\xi\sim N(m,0))$, 如果 $\mathbf{P}\{\xi=m\}=1$.

不过可以给出另一定义, 使之既包含*非退化* $(\sigma^2>0)$ 情形又包含*退化* $(\sigma^2=0)$ 情形. 为此, 考虑特征函数 $\varphi_\xi(t)\equiv\mathbf{E}\mathrm{e}^{it\xi},t\in\mathbb{R}$.

如果 $\mathbf{P}\{\xi=m\}=1$, 则显然

$$\varphi_\xi(t)=\mathrm{e}^{itm},\tag{2}$$

如果 $\xi\backsim N(m,\sigma^2),\sigma^2>0$, 则根据 §12(9) 式, 有

$$\varphi_\xi(t)=\mathrm{e}^{itm-\frac{t^2\sigma^2}{2}}.\tag{3}$$

易见, 对于 $\sigma^2=0$, (3) 式的右侧与 (2) 式的右侧相同. 由此及 §12 的定理 1, 可见参数为 m 和 $\sigma^2(|m|<\infty,\sigma^2\geqslant 0)$ 的高斯随机变量, 可以定义为特征函数 $\varphi_\xi(t)$ 由 (3) 式定义的随机变量. 运用特征函数的方法, 在多维情形特别方便.

设 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 是随机向量, 而

$$\varphi_\xi(\boldsymbol{t})=\mathbf{E}\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{\xi})},\boldsymbol{t}=(t_1,\cdots,t_n)\in\mathbb{R}^n,\tag{4}$$

是其特征函数 (见 §12 定义 2).

定义 1 称随机向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 服从*高斯分布*或*正态分布*, 如果其特征函数 $\varphi_\xi(\boldsymbol{t})$ 有如下形式:

$$\varphi_\xi(\boldsymbol{t})=\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{m})-\frac{1}{2}(\boldsymbol{R}\boldsymbol{t},\boldsymbol{t})},\tag{5}$$

其中 $\boldsymbol{m}=(m_1,\cdots,m_n),|m_k|<\infty$, 而 $\boldsymbol{R}=(r_{ij})$ 是 $n\times n$ 阶非负定矩阵 (简记为 $\boldsymbol{\xi}\sim N(\boldsymbol{m},\boldsymbol{R})$).

鉴于所引进的定义, 首先产生一个问题, (5) 式的函数是否特征函数? 现在证明它确实是特征函数.

为此, 首先假设 $\boldsymbol{R}$ 是非退化的. 那么存在逆矩阵 $\boldsymbol{A}=\boldsymbol{R}^{-1}$, 且函数

$$f(\boldsymbol{x})=\frac{|\boldsymbol{A}|^{1/2}}{(2\pi)^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{m}),(\boldsymbol{x}-\boldsymbol{m}))\right\}, \tag{6}$$

其中 $\boldsymbol{x}=(x_1,\cdots,x_n), |\boldsymbol{A}|=\det\boldsymbol{A}$. 该函数是非负的. 证明

$$\int_{\mathbb{R}^n}\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{x})}f(x)\mathrm{d}x=\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{m})-\frac{1}{2}(\boldsymbol{Rt},\boldsymbol{t})},$$

或同样地

$$I_n\equiv\int_{\mathbb{R}^n}\mathrm{e}^{i(\boldsymbol{t},\boldsymbol{x}-\boldsymbol{m})}\frac{|\boldsymbol{A}|^{1/2}}{(2\pi)^{n/2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{m}),(\boldsymbol{x}-\boldsymbol{m}))}\mathrm{d}x=\mathrm{e}^{-\frac{1}{2}(\boldsymbol{Rt},\boldsymbol{t})}. \tag{7}$$

在积分中作变量替换

$$\boldsymbol{x}-\boldsymbol{m}=\boldsymbol{Zu},\quad \boldsymbol{t}=\boldsymbol{Zv},$$

其中 $\boldsymbol{Z}$ 是正交矩阵, 满足

$$\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{RZ}=\boldsymbol{D},$$

而

$$\boldsymbol{D}=\begin{pmatrix}d_1 & & 0\\ & \ddots & \\ 0 & & d_n\end{pmatrix}$$

是对角矩阵, 其中 $d_i\geqslant 0$ (见 §8 引理的证明). 由于 $|\boldsymbol{R}|=\det\boldsymbol{R}\neq\boldsymbol{0}$, 可见 $d_i>0, i=1,\cdots,n$. 因此

$$|\boldsymbol{A}|=|\boldsymbol{R}^{-1}|=d_1^{-1}\cdots d_n^{-1}. \tag{8}$$

其次 (见 §12 第 1 小节的记号), 有

$$\begin{aligned}&i(\boldsymbol{t},\boldsymbol{x}-\boldsymbol{m})-\frac{1}{2}(\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{m}),(\boldsymbol{x}-\boldsymbol{m}))\\ &=i(\boldsymbol{Zv},\boldsymbol{Zu})-\frac{1}{2}(\boldsymbol{AZu},\boldsymbol{Zu})=i(\boldsymbol{Zv})^{\mathrm{T}}(\boldsymbol{Zu})-\frac{1}{2}(\boldsymbol{Zu})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{Zu})\\ &=i\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{AZu}=i\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{u}.\end{aligned}$$

由 (8) 式和 §12 的 (9) 式, 可得

$$\begin{aligned}I_n&=\frac{1}{(2\pi)^{n/2}(d_1\cdots d_n)^{1/2}}\int_{\mathbb{R}^n}\mathrm{e}^{i\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{u}}\mathrm{d}\boldsymbol{u}\\ &=\prod_{k=1}^{n}\frac{1}{(2\pi d_k)^{1/2}}\int_{-\infty}^{\infty}\mathrm{e}^{iv_ku_k-\frac{u_k^2}{2d_k}}\mathrm{d}u_k=\prod_{k=1}^{n}\mathrm{e}^{-\frac{v_k^2d_k}{2}}\\ &=\mathrm{e}^{-\frac{1}{2}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{Dv}}=\mathrm{e}^{-\frac{1}{2}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{RZv}}=\mathrm{e}^{-\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{Rt}}=\mathrm{e}^{-\frac{1}{2}(\boldsymbol{Rt},\boldsymbol{t})}.\end{aligned}$$

由 (6) 式亦可见

$$\int_{\mathbb{R}^n} f(x)\mathrm{d}x = 1. \tag{9}$$

于是, 函数 (5) 是 n 维 (非退化) 高斯分布的特征函数 (§3 第 3 小节).

现在, 设 $\boldsymbol{R}$ 是退化矩阵. 设 $\varepsilon > 0$, 考虑退化矩阵 $\boldsymbol{R}^\varepsilon \equiv \boldsymbol{R} + \varepsilon\boldsymbol{E}$, 其中 $\boldsymbol{E}$ 是单位矩阵. 那么, 由已证明的, 可见函数

$$\varphi^\varepsilon(\boldsymbol{t}) = \exp\left\{i(\boldsymbol{t},\boldsymbol{m}) - \frac{1}{2}(\boldsymbol{R}^\varepsilon\boldsymbol{t},\boldsymbol{t})\right\}$$

是特征函数:

$$\varphi^\varepsilon(\boldsymbol{t}) = \int_{\mathbb{R}^n} \mathrm{e}^{i(\boldsymbol{t},\boldsymbol{x})}\mathrm{d}F_\varepsilon(\boldsymbol{x}),$$

其中 $F_\varepsilon(\boldsymbol{x}) = F_\varepsilon(x_1,\cdots,x_n)$ 是 n 维分布函数.

当 $\varepsilon \to 0$ 时

$$\varphi^\varepsilon(\boldsymbol{t}) \to \varphi(\boldsymbol{t}) = \exp\left\{i(\boldsymbol{t},\boldsymbol{m}) - \frac{1}{2}(\boldsymbol{R}^\varepsilon\boldsymbol{t},\boldsymbol{t})\right\}.$$

极限函数 $\varphi(\boldsymbol{t})$ 在 0 点 $(0,\cdots,0)$ 连续. 因此, 根据定理 1 和第三章 §3 的练习题 1, $\varphi(\boldsymbol{t})$ 是特征函数.

于是, 定义 1 的适定性得证.

3. 高斯系的均值向量和协方差矩阵 现在说明特征函数的 (5) 式中, 向量 $\boldsymbol{m}$ 和矩阵 $\boldsymbol{R} = (r_{ij})$ 的含义.

由于

$$\ln\varphi_\xi(\boldsymbol{t}) = i(\boldsymbol{t},\boldsymbol{m}) - \frac{1}{2}(\boldsymbol{R}\boldsymbol{t},\boldsymbol{t}) = i\sum_{k=1}^n t_k m_k - \frac{1}{2}\sum_{k,l=1}^n r_{kl}t_k t_l, \tag{10}$$

由 §12 的 (35) 式, 以及矩和半不变量的换算公式, 可见

$$m_1 = s_\xi^{(1,0,\cdots,0)} = \mathbf{E}\xi_1, \cdots, m_n = s_\xi^{(0,\cdots,0,1)} = \mathbf{E}\xi_n.$$

类似地, 有

$$r_{11} = s_\xi^{(2,0,\cdots,0)} = \mathbf{D}\xi_1, \quad r_{12} = s_\xi^{(1,1,\cdots,0)} = \mathrm{cov}(\xi_1,\xi_2),$$

一般

$$r_{ij} = \mathrm{cov}(\xi_i,\xi_j).$$

于是, $\boldsymbol{m}$ 是 ξ 的*均值向量*, 而 $\boldsymbol{R}$ 是*协方差矩阵*.

如果矩阵 $\boldsymbol{R}$ 非退化, 则可以通过其他途径得到这些结果. 具体地说, 这时向量有 (6) 式给出的密度 $f(x)$, 而且通过直接推算可得

$$\begin{aligned} \mathbf{E}\xi_k &\equiv \int_{-\infty}^{\infty} x_k f(x)\mathrm{d}x = m_k, \\ \mathrm{cov}(\xi_k,\xi_l) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_k - m_k)(x_l - m_l)f(x_k,x_l)\mathrm{d}x_k\mathrm{d}x_l = r_{kl}. \end{aligned} \tag{11}$$

4. 高斯向量的性质　我们现在讨论高斯向量的若干性质.

定理 1　a) 对于高斯向量, 其分量不相关与分量独立等价.

b) 向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 是高斯向量, 当且仅当对于任意向量 $\boldsymbol{\lambda}=(\lambda_1,\cdots,\lambda_n)$, $\lambda_k\in\mathbb{R}$, 随机变量 $(\boldsymbol{\xi},\boldsymbol{\lambda})=\lambda_1\xi_1+\cdots+\lambda_n\xi_n$ 服从高斯分布.

证明　a) 如果向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 的分量不相关, 则由其特征函数 $\varphi_{\xi}(\boldsymbol{t})$ 的形式可见, 它等于各分量特征函数的乘积:

$$\varphi_{\xi}(\boldsymbol{t})=\prod_{k=1}^{n}\varphi_{\xi_k}(t_k).$$

所以, 由 §12 的定理 4 知, 分量 $\xi_1,\cdots,\xi_n$ 独立.

由于由独立性总是可以得出不相关性, 故逆命题显然成立.

b) 如果 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 是高斯向量, 则对于 $t\in\mathbb{R}$ (见 (5) 式)

$$\mathbf{E}\exp\{it(\lambda_1\xi_1+\cdots+\lambda_n\xi_n)\}=\exp\left\{it(\sum\lambda_k m_k)-\frac{t^2}{2}(\sum r_{kl}\lambda_k\lambda_l)\right\},$$

从而

$$(\boldsymbol{\xi},\boldsymbol{\lambda})\sim N(\sum\lambda_k m_k,\sum r_{kl}\lambda_k\lambda_l).$$

相反, 随机变量 $(\boldsymbol{\xi},\boldsymbol{\lambda})=\lambda_1\xi_1+\cdots+\lambda_n\xi_n$ 服从高斯分布, 则特别

$$\mathbf{E}e^{i(\boldsymbol{\xi},\boldsymbol{\lambda})}=e^{i\mathbf{E}(\boldsymbol{\xi},\boldsymbol{\lambda})-\frac{\mathbf{D}(\boldsymbol{\xi},\boldsymbol{\lambda})}{2}}=\exp\left\{i\sum\lambda_k\mathbf{E}\boldsymbol{\xi}_k-\frac{1}{2}\sum\lambda_k\lambda_l\mathrm{cov}(\xi_k,\xi_l)\right\}.$$

由于 $\lambda_1,\cdots,\lambda_n$ 的任意性, 可见 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 是高斯向量 (见定义 1).　□

注　设 $(\boldsymbol{\theta},\boldsymbol{\xi})$ 是高斯向量, 其中 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k),\boldsymbol{\xi}=(\xi_1,\cdots,\xi_l)$. 如果向量 $\boldsymbol{\theta}$ 和 $\boldsymbol{\xi}$ 不相关, 即 $\mathrm{cov}(\theta_i,\xi_j)=0,i=1,\cdots,k,j=1,\cdots,l$, 则它们就独立.

证明方法与定理 1 的命题 a)相同.

设 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 是高斯向量, 为简便计, 假设其均值向量为 0. 如果 $\mathrm{rang}\,\boldsymbol{R}=r<n$, 则由 §11 可见, 在变量 $\xi_1,\cdots,\xi_n$ 之间存在恰好 $n-r$ 个线性关系. 这时, 例如可以认为 $\xi_1,\cdots,\xi_r$ 线性无关, 而其余一切变量都可以通过它们线性表示. 因此向量 $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)$ 的所有基本性质都决定于前 r 个分量 $(\xi_1,\cdots,\xi_r)$. 而且它的协方差矩阵已经是退化的.

这样, 可以认为原向量的分量已经是线性无关的, 因而 $|\boldsymbol{R}|>0$.

设 $\boldsymbol{Z}$ 是把 $\boldsymbol{R}$ 变为对角矩阵的正交矩阵:

$$\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{Z}=\boldsymbol{D}.$$

如第 3 小节所指出的, 矩阵的所有对角元素都是正数, 从而有逆矩阵. 记 $B^2=\boldsymbol{D}$, 且设

$$\boldsymbol{\beta}=\boldsymbol{B}^{-1}\boldsymbol{Z}\boldsymbol{\xi}.$$

那么, 容易证明

$$\mathbf{E}e^{i(\boldsymbol{t},\boldsymbol{\beta})} = \mathbf{E}e^{i\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{t}} = e^{-\frac{1}{2}(E\boldsymbol{t},\boldsymbol{t})},$$

即向量 $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_n)$ 是具有不相关分量的高斯向量, 从而根据定理 1 是具有独立分量的高斯向量. 若记 $\boldsymbol{A} = \boldsymbol{ZB}$, 则原高斯向量 $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)$ 可以表示为

$$\boldsymbol{\xi} = \boldsymbol{A\beta}, \tag{12}$$

其中 $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_n)$ 是具有独立分量的高斯向量, 且 $\beta_k \sim N(0,1)$. 由此得如下结果: 设 $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)$ 是具有线性不相关分量的向量, $\mathbf{E}\xi_k = 0, k = 1, \cdots, n$, 则此向量是高斯向量的充分和必要条件是, 存在独立高斯变量 $\beta_1, \cdots, \beta_n, \beta_k \sim N(0,1)$, 和 n 阶非退化矩阵 $\boldsymbol{A}$, 使 $\boldsymbol{\xi} = \boldsymbol{A\beta}$. 这时 $\boldsymbol{R} = \boldsymbol{AA}^{\mathrm{T}}$ 是向量 ξ 的协方差矩阵.

如果 $\boldsymbol{R} \neq 0$, 则根据克拉默 – 施密特正交化方法 (见 §11), 有

$$\xi_k = \widehat{\xi}_k + b_k \varepsilon_k, \quad k = 1, \cdots, n, \tag{13}$$

其中由于高斯性, 向量 $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n) \sim N(\mathbf{0}, E)$,

$$\widehat{\xi}_k = \sum_{l=1}^{k-1} (\xi_k, \varepsilon_l)\varepsilon_l, \tag{14}$$

$$b_k = \|\xi_k - \widehat{\xi}_k\|, \tag{15}$$

且

$$\mathscr{L}\{\xi_1, \cdots, \xi_k\} = \mathscr{L}\{\varepsilon_1, \cdots, \varepsilon_k\}. \tag{16}$$

由正交展开式 (13), 立即得

$$\widehat{\xi}_k = \mathbf{E}(\xi_k | \xi_{k-1}, \cdots, \xi_1). \tag{17}$$

鉴于 (16) 和 (14) 式, 由此可见对于高斯变量, 条件数学期望 $\mathbf{E}(\xi_k | \xi_{k-1}, \cdots, \xi_1)$ 是 $\xi_1, \cdots, \xi_{k-1}$ **的线性函数**:

$$\mathbf{E}(\xi_k | \xi_{k-1}, \cdots, \xi_1) = \sum_{i=1}^{k-1} a_i \xi_i. \tag{18}$$

(对于 $k = 2$ 的情形, 在 §8 曾得到该结果.)

由于根据 §8 中对定理 1 的注知, $\mathbf{E}(\xi_k | \xi_{k-1}, \cdots, \xi_1)$ (均方意义上) 是由 $\xi_1, \cdots, \xi_{k-1}$ 对 ξ_k 的最优估计量, 则由 (18) 式可见, 对于高斯分布, **最优估计量是线性的**.

对于 $(\boldsymbol{\theta}, \boldsymbol{\xi})$ 是高斯向量的情形, 我们利用以上的结果由向量 $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)$, 来求向量 $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$ 的最优估计. 记

$$\boldsymbol{m}_{\boldsymbol{\theta}} = \mathbf{E}\boldsymbol{\theta}, \quad \boldsymbol{m}_{\boldsymbol{\xi}} = \mathbf{E}\boldsymbol{\xi}$$

为均值列向量, 而

$$\mathbf{D}_{\theta\theta} \equiv \mathrm{cov}(\boldsymbol{\theta}, \boldsymbol{\theta}) \equiv (\mathrm{cov}(\theta_i, \theta_j)), 1 \leqslant i, j \leqslant k,$$
$$\mathbf{D}_{\theta\xi} \equiv \mathrm{cov}(\boldsymbol{\theta}, \boldsymbol{\xi}) \equiv (\mathrm{cov}(\theta_i, \xi_j)), 1 \leqslant i \leqslant k, 1 \leqslant j \leqslant l,$$
$$\mathbf{D}_{\xi\xi} \equiv \mathrm{cov}(\boldsymbol{\xi}, \boldsymbol{\xi}) \equiv (\mathrm{cov}(\xi_i, \xi_j)), 1 \leqslant i, j \leqslant l,$$

是协方差矩阵. 假设矩阵 $\mathbf{D}_{\xi\xi}$ 有逆矩阵. 那么, 下列定理成立 (对照 §8 定理 2.)

定理 2 (正态相关性定理, 向量情形)　*对于正态随机向量 $(\boldsymbol{\theta}, \boldsymbol{\xi})$, 由 $\boldsymbol{\xi}$ 对向量 $\boldsymbol{\theta}$ 的最优估计量 $\mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi})$, 及其误差矩阵*

$$\Delta = \mathbf{E}[\boldsymbol{\theta} - \mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi})][\boldsymbol{\theta} - \mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi})]^{\mathrm{T}},$$

分别由下面的公式给出:

$$\mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi}) = \boldsymbol{m}_{\theta} + \mathbf{D}_{\theta\xi}\mathbf{D}_{\xi\xi}^{-1}(\boldsymbol{\xi} - \boldsymbol{m}_{\xi}), \tag{19}$$
$$\Delta = \mathbf{D}_{\theta\theta} - \mathbf{D}_{\theta\xi}\mathbf{D}_{\xi\xi}^{-1}\mathbf{D}_{\theta\xi}^{\mathrm{T}}. \tag{20}$$

证明　记向量

$$\boldsymbol{\eta} = (\boldsymbol{\theta} - \boldsymbol{m}_{\theta}) - \mathbf{D}_{\theta\xi}\mathbf{D}_{\xi\xi}^{-1}(\boldsymbol{\xi} - \boldsymbol{m}_{\xi}). \tag{21}$$

那么, 可以直接验证 $\mathbf{E}\boldsymbol{\eta}(\boldsymbol{\xi} - \boldsymbol{m}_{\xi})^{\mathrm{T}} = 0$, 即向量 $\boldsymbol{\eta}$ 与向量 $\boldsymbol{\xi} - \boldsymbol{m}_{\xi}$ 不相关. 由于向量 $(\boldsymbol{\theta}, \boldsymbol{\xi})$ 的高斯性, 则 $(\boldsymbol{\eta}, \boldsymbol{\xi})$ 也是高斯向量. 由此以及定理 1 的注, 知向量 $\boldsymbol{\eta}$ 与 $\boldsymbol{\xi} - \boldsymbol{m}_{\xi}$ 独立. 因此, 向量 $\boldsymbol{\eta}$ 与 $\boldsymbol{\xi}$ 独立, 从而 $\mathbf{E}(\boldsymbol{\eta}|\boldsymbol{\xi}) = \mathbf{E}\boldsymbol{\eta} = 0$. 所以

$$\mathbf{E}[\boldsymbol{\theta} - \boldsymbol{m}_{\theta}|\boldsymbol{\xi}] - \mathbf{D}_{\theta\xi}\mathbf{D}_{\xi\xi}^{-1}(\boldsymbol{\xi} - \boldsymbol{m}_{\xi}) = 0,$$

于是, 公式 (19) 得证.

为证明 (20) 式, 考虑条件协方差

$$\mathrm{cov}(\boldsymbol{\theta}, \boldsymbol{\theta}|\boldsymbol{\xi}) = \mathbf{E}\{[\boldsymbol{\theta} - \mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi})][\boldsymbol{\theta} - \mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi})]^{\mathrm{T}}|\boldsymbol{\xi}\}. \tag{22}$$

由 (19) 和 (21) 式可见 $\boldsymbol{\theta} - \mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi}) = \boldsymbol{\eta}$, 故由 $\boldsymbol{\eta}$ 与 $\boldsymbol{\xi}$ 独立, 可见

$$\begin{aligned}
\mathrm{cov}(\boldsymbol{\theta}, \boldsymbol{\theta}|\boldsymbol{\xi}) &= \mathbf{E}(\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}|\boldsymbol{\xi}) = \mathbf{E}\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}} \\
&= \mathbf{D}_{\theta\theta} + \mathbf{D}_{\theta\xi}\mathbf{D}_{\xi\xi}^{-1}\mathbf{D}_{\xi\xi}\mathbf{D}_{\xi\xi}^{-1}\mathbf{D}_{\theta\xi}^{\mathrm{T}} - 2\mathbf{D}_{\theta\xi}\mathbf{D}_{\xi\xi}^{-1}\mathbf{D}_{\xi\xi}\mathbf{D}_{\xi\xi}^{-1}\mathbf{D}_{\theta\xi}^{\mathrm{T}} \\
&= \mathbf{D}_{\theta\theta} - \mathbf{D}_{\theta\xi}\mathbf{D}_{\xi\xi}^{-1}\mathbf{D}_{\theta\xi}^{\mathrm{T}}.
\end{aligned}$$

因为 $\mathrm{cov}(\boldsymbol{\theta}, \boldsymbol{\theta}|\boldsymbol{\xi})$ 与 "偶然性" 无关, 所以

$$\Delta = \mathbf{E}\,\mathrm{cov}(\boldsymbol{\theta}, \boldsymbol{\theta}|\boldsymbol{\xi}) = \mathrm{cov}(\boldsymbol{\theta}, \boldsymbol{\theta}|\boldsymbol{\xi}),$$

于是, (20) 式得证.　□

系 设 $(\theta,\xi_1,\cdots,\xi_n)$ 是 $n+1$ 维高斯向量, 而且 $\xi_1,\cdots,\xi_n$ 独立. 那么,

$$\mathbf{E}(\theta|\xi_1,\cdots,\xi_n)=\mathbf{E}\theta+\sum_{i=1}^{n}\frac{\operatorname{cov}(\theta,\xi_i)}{\mathbf{D}\xi_i}(\xi_i-\mathbf{E}\xi_i),$$

$$\Delta=\mathbf{D}\theta-\sum_{i=1}^{n}\frac{\operatorname{cov}^2(\theta,\xi_i)}{\mathbf{D}\xi_i}$$

(对照 §8 公式 (12),(13)).

5. 高斯向量的线性流形的封闭性 设 $\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\cdots$ 是依概率收敛于向量 $\boldsymbol{\xi}$ 的高斯随机向量序列. 现在证明向量 $\boldsymbol{\xi}$ 仍然服从高斯分布.

由于定理 1 的命题 a), 只需对随机变量证明这一事实.

设 $m_n=\mathbf{E}\xi_n,\sigma_n^2=\mathbf{D}\xi_n$. 那么, 根据勒贝格控制收敛定理, 有

$$\lim_{n\to\infty}\mathrm{e}^{itm_n-\frac{1}{2}\sigma_n^2t^2}=\lim_{n\to\infty}\mathbf{E}\mathrm{e}^{it\xi_n}=\mathbf{E}\mathrm{e}^{it\xi}.$$

由于左侧极限的存在性, 可见存在 m_n 和 σ_n^2, 使

$$m=\lim_{n\to\infty}m_n,\quad \sigma^2=\lim_{n\to\infty}\sigma_n^2.$$

从而,

$$\mathbf{E}\mathrm{e}^{it\xi}=\mathrm{e}^{itm-\frac{1}{2}\sigma^2t^2},$$

即 $\xi\sim N(m,\sigma^2)$.

特别, 由此可见, 由高斯随机变量 $\xi_1,\xi_2,\cdots$ 生成的封闭线性流形 $\overline{\mathscr{L}}(\xi_1,\xi_2,\cdots)$ (见 §11 第 5 小节), 仍然由高斯随机变量生成.

6. 一般高斯系及其性质 现在, 讨论一般高斯系的定义.

定义 2 设 $\mathfrak{A}$ 是某一下标的集合. 随机变量 $\xi=\{\xi_\alpha\},\alpha\in\mathfrak{A}$ 的全体称为**高斯系**, 如果对于任意 $\alpha_1,\cdots,\alpha_n\in\mathfrak{A},n\geqslant1$, 随机向量 $(\xi_{\alpha_1},\cdots,\xi_{\alpha_n})$ 是高斯向量.

现在指出高斯系的一些性质.

a) 如果 $\xi=\{\xi_\alpha\},\alpha\in\mathfrak{A}$ 是高斯系, 则任意 $\xi'=\{\xi'_{\alpha'}\},\alpha'\in\mathfrak{A}'\subset\mathfrak{A}$ 子系也是高斯系.

b) 如果 $\xi_\alpha,\alpha\in\mathfrak{A}$ 是独立高斯变量, 则 $\xi=\{\xi_\alpha\},\alpha\in\mathfrak{A}$ 是高斯系.

c) 如果 $\xi=\{\xi_\alpha\},\alpha\in\mathfrak{A}$ 是高斯系, 则包含由形如 $\sum_{i=1}^n c_{\alpha_i}\xi_{\alpha_i}$ 的变量及其均方极限的闭线性流形 $\overline{\mathscr{L}}(\xi)$ 是高斯系.

注意, 性质 a) 的逆命题一般不成立. 例如, 假设 ξ_1 和 η_1 独立, 且 $\xi_1\sim N(0,1)$, $\eta_1\sim N(0,1)$. 定义系统如下

$$(\xi,\eta)=\begin{cases}(\xi_1,|\eta_1|), & \text{若 } \xi_1\geqslant0,\\ (\xi_1,-|\eta_1|), & \text{若 } \xi_1<0.\end{cases}\tag{23}$$

那么, 不难验证 ξ 和 η 中每一个都是高斯变量, 然而 (ξ,η) 却不是高斯向量.

设 $\xi = \{\xi_\alpha\}, \alpha \in \mathfrak{U}$ 是高斯系, 其均值 “向量” 为 $m = \{m_\alpha\}, \alpha \in \mathfrak{U}$, 而协方差 “矩阵” 为 $\boldsymbol{R} = (r_{\alpha\beta})_{\alpha,\beta\in\mathfrak{U}}$, 其中 $m_\alpha \in \mathbf{E}\xi_\alpha$.“矩阵” $\boldsymbol{R}$ 显然是对称的 $(r_{\alpha\beta} = r_{\beta\alpha})$, 并在如下意义上是非负定的: 对于任意在 $\mathbb{R}^{\mathfrak{U}}$ 取值的 “向量” $c = \{c_\alpha\}_{\alpha\in\mathfrak{U}}$, 只有有限个坐标 c_α 不为 0,

$$(\boldsymbol{R}c, c) \equiv \sum_{\alpha,\beta} r_{\alpha\beta} c_\alpha c_\beta \geqslant 0. \tag{24}$$

现在提出反问题. 假设给定一参数的集合 $\mathfrak{U} = \{\alpha\}$、“向量” $m = \{m_\alpha\}_{\alpha\in\mathfrak{U}}$ 和对称非负定 “矩阵” $\boldsymbol{R} = (r_{\alpha\beta})_{\alpha,\beta\in\mathfrak{U}}$. 问是否存在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 和空间上的高斯随机变量系 $\xi = \{\xi_\alpha\}, \alpha \in \mathfrak{U}$, 使

$$\mathbf{E}\xi_\alpha = m_\alpha,$$

$$\operatorname{cov}(\xi_\alpha, \xi_\beta) = r_{\alpha\beta}, \quad \alpha, \beta \in \mathfrak{U}?$$

如果取有限组 $\alpha_1, \cdots, \alpha_n$, 则根据向量 $\overline{m} = (m_{\alpha_1}, \cdots, m_{\alpha_n})$ 和矩阵 $\overline{\boldsymbol{R}} = (r_{\alpha\beta})_{\alpha,\beta\in\mathfrak{U}}$, $\alpha, \beta = \alpha_1, \cdots, \alpha_n \in \mathbb{R}^n$, 可以建立高斯分布 $F_{\alpha_1,\cdots,\alpha_n}(x_1, \cdots, x_n)$, 使其特征函数为

$$\varphi(\boldsymbol{t}) = \mathrm{e}^{i(t,\bar{m}) - \frac{1}{2}(\overline{\boldsymbol{R}}\boldsymbol{t},\boldsymbol{t})}, \quad \boldsymbol{t} = (t_{\alpha_1}, \cdots, t_{\alpha_n}).$$

不难验证, 分布族

$$\{F_{\alpha_1,\cdots,\alpha_n}(x_1, \cdots, x_n); \alpha_i \in \mathfrak{U}\}$$

是一致的. 从而, 根据柯尔莫戈洛夫定理 (§9 定理 1 及其注 2) 对于上面所提问题的答案是肯定的.

7. 高斯随机序列和高斯过程　假如 $\mathfrak{U} = \{1, 2, \cdots\}$, 则按 §5 所采用的术语, 随机变量系 $\boldsymbol{\xi} = \{\xi_\alpha\}, \alpha \in \mathfrak{U}$ 称做*随机序列*, 并记作 $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots)$. 高斯序列 $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots)$ 完全由均值向量 $\boldsymbol{m} = (m_1, m_2, \cdots)$ 和协方差矩阵 $\boldsymbol{R} = (r_{ij})$, 其中 $r_{ij} = \operatorname{cov}(\xi_i, \xi_j)$ 描绘. 特别, 如果 $r_{ij} = \sigma^2 \delta_{ij}$, 则 $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots)$ 是独立随机变量的高斯序列, 其中 $\xi_i \sim N(m_i, \sigma_i^2), i \geqslant 1$.

对于 $\mathfrak{U} = [0, 1], [0, \infty), (-\infty, \infty), \cdots$ 的情形, 随机变量系 $\xi = \{\xi_t\}, t \in \mathfrak{U}$ 称做*连续时间随机过程*.

现在举几个高斯随机过程的例子. 假如认为其均值为 0, 则这样过程的概率性质完全决定于其协方差矩阵 $\boldsymbol{R} = (r_{st}), s, t \in \mathfrak{U}$ 的形式. 我们用 $r(s, t)$ 表示 r_{st}, 并称此 s 和 t 函数为*协方差函数*.

例 1　如果 $\mathfrak{U} = [0, \infty)$, 而

$$r(s, t) = \min(s, t), \tag{25}$$

则以 $r(s, t)$ 为协方差函数 (见练习题 2), 且 $B_0 \equiv 0$ 的高斯过程 $B = \{B_t\}_{t\geqslant 0}$ 称做*布朗运动过程或维纳过程*.

注意, 该过程具有独立增量, 即对于任意 $t_1 < t_2 < \cdots < t_n$ 随机变量

$$B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}$$

是独立的. 事实上, 由于高斯性只需验证它们两两不相关. 实际上, 若 $s < t < u < v$, 则

$$\begin{aligned}&\mathbf{E}(B_t - B_s)(B_v - B_u)\\ &= [r(t,v) - r(t,u)] - [r(s,v) - r(s,u)]\\ &= (t-t) - (s-s) = 0.\end{aligned}$$

注 §9 第 4 小节考虑的更新过程的例子 (是根据独立随机变量序列 $\sigma_1, \sigma_2, \cdots$ 的构造性给出的), 使得可能类似地建立某种类型的布朗运动的想法产生了.

以独立同分布的标准高斯随机变量序列 $\xi_1, \xi_2, \cdots$ 为基础, 其中 $\xi_i \sim N(0,1)$, 来构造布朗运动的情形确实存在.

例如, 构造随机变量

$$B_t = \frac{\sqrt{2}}{\pi}\sum_{n=1}^{\infty}\frac{\xi_n}{n+1/2}\sin((n+1/2)\pi t), \quad t \in [0,1]. \tag{26}$$

由下面将给出的 "两级数" 定理 (第四章 §3 定理 2), 可见对于每一个 $t \in [0,1]$, 定义 B_t 的级数 $(\mathbf{P}-\text{a.c.})$ 收敛. 进一步更细致地讨论表明, 此级数 $(\mathbf{P}-\text{a.c.})$ 一致收敛, 从而过程 $B = (B_t)_{0\leqslant t\leqslant 1}(\mathbf{P}-\text{a.c.})$ 有连续轨道. 该过程的正态性, 由定理 1 的 b) 和第 5 小节中关于 "高斯随机变量依概率收敛的极限, 仍然服从高斯分布" 的论断可以证明. 同样不难证明, 协方差函数为 $r(s,t) = \mathbf{E}B_sB_t = \min(s,t)$.

这样, 在 (26) 式中建立的过程 $B = (B_t)_{0\leqslant t\leqslant 1}$ 满足布朗运动过程定义中的所有要求, 此外该过程 $(\mathbf{P}-\text{a.c.})$ 有连续轨道. 通常, (物理应用所希望的和得到证实的) 轨道的连续性, 包含在布朗运动的定义之中. 我们看到, 这样的过程确实存在.

我们再指出一个建立布朗运动过程的熟知方法, 它基于在 §11 第 5 小节中引进的哈尔函数 $H_n(x), x \in [0,1], n = 1,2,\cdots$

由哈尔函数 $H_n(x)$ 建立绍德尔 (A. Schauder) 函数 $S_n(t), t \in [0,1], n = 1,2,\cdots$,

$$S_n(t) = \int_0^t H_n(x)\mathrm{d}x. \tag{27}$$

那么, 如果 $\xi_0, \xi_1, \xi_2, \cdots$ 是独立同标准高斯分布随机变量序列, 其中 $\xi_i \sim N(0,1)$, 则级数对于 $t \in [0,1]$,

$$B_t = \sum_{n=1}^{\infty}\xi_n S_n(t) \tag{28}$$

以概率 1 一致收敛. 过程 $B = (B_t)_{0\leqslant t\leqslant 1}$ 是布朗运动.

例 2 过程 $B^0 = (B_t^0), t \in \mathfrak{U} = [0,1], B_0^0 \equiv 0$ 和

$$r(s,t) = \min(s,t) - st, \tag{29}$$

称做条件维纳过程或布朗桥 (注意, 由于 $r(1,1) = 0$, 故 $\mathbf{P}\{B_1^0 = 0\} = 1$).

例 3 过程 $X = (X_t), t \in \mathfrak{U} = (-\infty, \infty)$ 和

$$r(s,t) = \mathrm{e}^{-|t-s|} \tag{30}$$

称做高斯 – 马尔可夫过程.

8. 布朗运动的一个简单例子 我们介绍一个布朗运动的有趣性质, 其证明是 §10 中博雷尔 – 坎泰利引理应用的很好的演示 (确切地说该引理的系 1).

定理 3 设 $B = (B_t)_{t \geqslant 0}$ 是标准布朗运动. 那么, 对于任意 $T > 0$, 依概率 1

$$\lim_{n \to \infty} \sum_{k=1}^{2^n T} [B_{k2^{-n}} - B_{(k-1)2^{-n}}]^2 = T. \tag{31}$$

证明 不失普遍性, 可以认为 $T = 1$. 设

$$A_n^\varepsilon = \left\{ \omega : \left| \sum_{k=1}^{2^n} [B_{k2^{-n}} - B_{(k-1)2^{-n}}]^2 - 1 \right| \geqslant \varepsilon \right\}.$$

由于 $B_{k2^{-n}} - B_{(k-1)2^{-n}}$ 是高斯随机变量, 且均值为 0, 而方差等于 2^{-n}, 可见

$$\mathbf{D} \left(\sum_{k=1}^{2^n} [B_{k2^{-n}} - B_{(k-1)2^{-n}}]^2 \right) = 2^{-n+1};$$

因此, 由切比雪夫不等式, 有 $\mathbf{P}(A_n^\varepsilon) \leqslant \varepsilon^{-2} 2^{-n+1}$. 从而

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n^\varepsilon) \leqslant \varepsilon^{-2} \sum_{n=1}^{\infty} 2^{-n+1} = 2\varepsilon^{-2} < \infty. \tag{32}$$

于是, 由此估计式和博雷尔 – 坎泰利引理的系 1 (§10), 得所要求的关系式 (31). □

9. 练习题

1. 设 ξ_1, ξ_2, ξ_3 是独立高斯随机变量, $\xi_i \sim N(0,1)$. 证明

$$\frac{\xi_1 + \xi_2 \xi_3}{\sqrt{1 + \xi_3^2}} \sim N(0,1).$$

(由此产生了对于有趣的研究课题: 描绘独立高斯随机变量 $\xi_1, \cdots, \xi_n$ 的非线性变换, 使变换的结果仍然是高斯随机变量.)

2. 证明, 由 (25),(29) 和 (30) 各式的函数 $r(s,t)$ 给出的 “矩阵” $\boldsymbol{R} = (r(s,t))_{s,t \in \mathfrak{U}}$ 是非负定的.

3. 设 $\boldsymbol{A}$ 是 $m\times n$ 阶矩阵. 称 $\boldsymbol{A}^{\otimes}$ 为 $\boldsymbol{A}$ 的 $n\times m$ 阶伪逆矩阵, 如果存在矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$, 使

$$\boldsymbol{A}\boldsymbol{A}^{\otimes}\boldsymbol{A}=\boldsymbol{A},\quad \boldsymbol{A}^{\otimes}=\boldsymbol{U}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}.$$

证明, 由这些条件决定的矩阵 $\boldsymbol{A}^{\otimes}$ 存在而且唯一.

4. 证明在正态相关定理的公式 (19) 和 (20) 中, 对于 $\mathbf{D}_{\xi\xi}$ 是退化矩阵的情形, 如果在这些公式中用伪逆矩阵 $\mathbf{D}_{\xi\xi}^{\oplus}$ 代替 $\mathbf{D}_{\xi\xi}^{-1}$, 则正态相关定理仍然成立.

5. 设 $(\boldsymbol{\theta},\boldsymbol{\xi})=(\theta_1,\cdots,\theta_k;\xi_1,\cdots,\xi_l)$ 是具有非退化矩阵 $\Delta\equiv\mathbf{D}_{\theta\theta}-\mathbf{D}_{\xi\xi}^{\oplus}\mathbf{D}_{\theta\xi}^{\mathrm{T}}$ 的高斯向量. 证明分布函数 $\mathbf{P}\{\boldsymbol{\theta}\leqslant\boldsymbol{a}|\boldsymbol{\xi}\}=\mathbf{P}\{\theta_1\leqslant a_1,\cdots,\theta_k\leqslant a_k|\boldsymbol{\xi}\}(\mathbf{P}-\text{a.c.})$ 有密度 $p\{a_1,\cdots,a_k|\boldsymbol{\xi}\}$, 且

$$p\{a_1,\cdots,a_k|\boldsymbol{\xi}\}=\frac{|\Delta|^{-1/2}}{(2\pi)^{k/2}}\exp\left\{-\frac{1}{2}[\boldsymbol{a}-\mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi})]^{\mathrm{T}}\Delta^{-1}[\boldsymbol{a}-\mathbf{E}(\boldsymbol{\theta}|\boldsymbol{\xi})]\right\}.$$

6. (C.H. 伯恩斯坦) 设 ξ 和 η 是具有有限方差的独立同分布随机变量. 证明, 如果 $\xi+\eta$ 和 $\xi-\eta$ 独立, 则 ξ 和 η 是高斯随机变量.

7. **默瑟** (J. Mercer) **定理**. 假设 $r(s,t)$ 是 $[a,b]\times[a,b](-\infty<a<b<\infty)$ 上连续的协方差函数. 证明, 对于无限多个 $\lambda_k>0$, 方程

$$\lambda\int_a^b r(s,t)u(t)\mathrm{d}t=u(s),a\leqslant s\leqslant b,$$

相应的连续解 $\{u_k,k\geqslant1\}$ 的解系, 构成 $L^2(a,b)$ 中的完全规范正交系, 使

$$r(s,t)=\sum_{k=1}^{\infty}\frac{u_k(s)u_k(t)}{\lambda_k},$$

其中级数在 $[a,b]\times[a,b]$ 上绝对收敛且一致收敛.

8. 设 $X=\{X_t,t\geqslant0\}$ 是高斯过程, 其中 $\mathbf{E}X_t=0$, 而协方差函数 $r(s,t)=\mathrm{e}^{-|t-s|},s,t>0$. 设 $0<t_1<\cdots<t_n$, 而 $f_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$ 是随机变量 $X_{t_1},\cdots,X_{t_n}$ 的密度. 证明:

$$f_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=\left[(2\pi)^n\prod_{i=2}^{n}(1-\mathrm{e}^{2(t_{i-1}-t_i)})\right]^{-1/2}\times\exp\left\{-\frac{x_1^2}{2}-\frac{1}{2}\sum_{i=2}^{n}\frac{[x_i-\mathrm{e}^{(t_{i-1}-t_i)}x_{i-1}]^2}{1-\mathrm{e}^{2(t_{i-1}-t_i)}}\right\}.$$

9. 设 $f=\{f_n,n\geqslant1\}\subset L^2(0,1)$ 是完备规范正交系, 而 $\{\xi_n\}$ 是独立同分布 $N(0,1)$ 随机变量. 证明过程

$$B_t=\sum_{n\geqslant1}\xi_n\int_0^t f_n(u)\mathrm{d}u$$

是布朗运动.

10. 证明当 $(\xi,\eta_1,\cdots,\eta_n)$ 高斯系时, 条件数学期望 $\mathbf{E}(\xi|\eta_1,\cdots,\eta_n)$ 等于广义数学期望 $\widehat{\mathbf{E}}(\xi|\eta_1,\cdots,\eta_n)$.

11. 设 $(\xi,\eta_1,\cdots,\eta_k)$ 是高斯系, 说明条件数学期望 $\mathbf{E}(\xi^n|\eta_1,\cdots,\eta_k), n\geqslant 1$ (作为 $\eta_1,\cdots,\eta_k$ 的函数) 的构造.

12. 设 $X=(X_k)_{1\leqslant k\leqslant n}$ 和 $Y=(Y_k)_{1\leqslant k\leqslant n}$ 是两个高斯随机序列, 且 $\mathbf{E}X_k=\mathbf{E}Y_k, \mathbf{D}X_k=\mathbf{D}Y_k, 1\leqslant k\leqslant n$, 而

$$\operatorname{cov}(X_k,X_l)\leqslant \operatorname{cov}(Y_k,Y_l), 1\leqslant k,l\leqslant n.$$

证明斯莱皮恩 (P. Slepian) 不等式: 对于任意 $x\in R$,

$$\mathbf{P}\left\{\sup_{1\leqslant k\leqslant n} X_k<x\right\}\leqslant \mathbf{P}\left\{\sup_{1\leqslant k\leqslant n} Y_k<x\right\}.$$

13. 证明, 如果 $B^\circ=(B_t^\circ)_{0\leqslant t\leqslant 1}$ 是布朗桥, 则过程 $B=(B_t)_{t\geqslant 0}$ 是布朗运动, 其中 $B_t=(1+t)B^\circ_{t/(1+t)}$.

14. 对于布朗运动 $B=(B_t)_{t\geqslant 0}$, 验证下列过程也是布朗运动:

$$\begin{aligned}
&B_t^{(1)}=-B_t;\\
&B_t^{(2)}=tB_{1/t}, t>0 \text{ 和 } B_0^{(2)}=0;\\
&B_t^{(3)}=B_{t+s}-B_s, s>0;\\
&B_t^{(4)}=B_T-B_{T-t}, 0\leqslant t\leqslant T, T>0;\\
&B_t^{(5)}=\frac{1}{a}B_{a^2t}, a>0\ (\text{自模态性}).
\end{aligned}$$

15. 设 $X=(X_k)_{1\leqslant k\leqslant n}$ 是高斯随机序列, 且

$$m=\max_{1\leqslant k\leqslant n}\mathbf{E}X_k,\quad \sigma^2=\max_{1\leqslant k\leqslant n}\mathbf{D}X_k,$$

而对于某个 a,

$$\mathbf{P}\left\{\sup_{1\leqslant k\leqslant n}(X_k-\mathbf{E}X_k)\geqslant a\right\}\leqslant\frac{1}{2}.$$

那么, 博雷尔不等式成立:

$$\mathbf{P}\left\{\sup_{1\leqslant k\leqslant n} X_k>x\right\}\leqslant 2\Psi\left(\frac{x-m-a}{\sigma}\right),$$

其中

$$\Psi(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty \mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y.$$

16. 设 (X,Y) 是二维高斯随机变量, $\mathbf{E}X=\mathbf{E}Y=0, \mathbf{E}X^2>0, \mathbf{E}Y^2>0$, 其相关系数为

$$\rho=\frac{\mathbf{E}XY}{\sqrt{\mathbf{E}X^2\mathbf{E}Y^2}}.$$

证明

$$\mathbf{P}\{XY<0\}=1-2\mathbf{P}\{X>0,Y>0\}=\pi^{-1}\arccos\rho.$$

17. 设 $Z=XY$, 其中 $X\sim N(0,1)$, 且 $\mathbf{P}\{Y=1\}=\mathbf{P}\{Y=-1\}=1/2$. 求随机向量 (X,Z) 和 (Y,Z) 的分布, 以及 $X+Z$ 的分布. 证明, $Z\sim N(0,1)$, 且 X 和 Z 不相关, 但是不独立.

18. 详细证明, 由 (26),(28) 式定义的过程 $(B_t)_{0\leqslant t\leqslant 1}$ 是布朗运动.

19. 设 $B^\mu=(B_t+\mu t)_{t\geqslant 0}$ 是带漂移的布朗运动.

(a) 求变量 $B_{t_1}^\mu+B_{t_2}^\mu, t_1<t_2$, 的分布;

(b) 对于 $t_0<t_1<t_2$, 求 $\mathbf{E}B_{t_0}^\mu B_{t_1}^\mu$ 和 $\mathbf{E}B_{t_0}^\mu B_{t_1}^\mu B_{t_2}^\mu$.

20. 对于上题的过程 B^μ, 对于 $t_1<t_2$ 和 $t_1>t_2$, 求条件分布:

$$\mathbf{P}\{B_{t_2}^\mu\in\cdot|B_{t_1}^\mu\};$$

对于 $t_0<t_1<t_2$, 求条件分布:

$$\mathbf{P}\{B_{t_2}^\mu\in\cdot|B_{t_0}^\mu,B_{t_1}^\mu\}.$$

第三章 概率测度的接近程度和收敛性. 中心极限定理

对于概率论教程形式上的结构, 极限定理是概率论初等章节的一种上层建筑, 其中所有问题都具有最终的、纯算术的特点. 不过, 概率论认识的价值, 实际上只有通过极限定理才揭示出来. 何况, 没有极限定理就无法理解我们的全部学科 —— 概率论有关概念的真实内容.

Б. В. 格涅坚科, А. Н. 柯尔莫戈洛夫.
《独立随机变量之和的极限分布》[16]

§1. 概率测度和分布的弱收敛

1. 概述 概率论许多结果都具有*极限定理*的形式. 以极限定理的形式, 建立了 J. 伯努利大数定律和棣莫弗 - 拉普拉斯定理, 可以说, 从而奠定了真正概率论的基础, 特别是指出了众多研究的方向, 说明了不同形式的 "大数定律" 和 "中心极限定理" 成立的条件. 以极限定理的形式还创立了, "在稀有事件情形下, 用 '泊松分布' 逼近二项分布的" *泊松定理*. 在这些定理的例子中, 以及在有关棣莫弗 - 拉普拉斯定理和泊松定理的收敛速度的结果中, 已经可以看到, 在概率论中必然涉及分布的不同形式. 而说明*收敛速度*, 需要引进分布之间各种接近程度的 "自然" 度量.

在这一章将讨论概率分布的*收敛性*, 及其*接近程度*的某些一般性问题. 这一节讨论度量空间中*概率测度弱收敛*的一般理论问题. (特别, 伯努利大数定律和棣莫弗 - 拉普拉斯定理 —— "中心极限定理" 的 "始祖", 正是属于这一范围.) 从 §3 开始就可以清楚地看到特征函数方法, 是证明 $\mathbb{R}^n$ 中概率分布弱收敛的极限定理最强有力的工具之一. 在 §7 中将讨论弱收敛的 "可度量化" 问题. 在 §9 中将讨论分布另一种形式的收敛性: *按变差收敛* (一种比弱收敛更强的收敛). 在中心极限定理和泊松定理中, 收敛速度最简单结果的证明将在 §11 和 §12 中给出. 在 §13 中将 §1 和 §2 中关于弱收敛的结果用于某些 (原则上重要的) 数理统计问题.

2. 伯努利概型中的收敛性 作为开始, 我们回忆*伯努利概型中大数定律*的提法 (第一章 §5).

设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{P}\{\xi_i = 1\} = p, \mathbf{P}\{\xi_i = 0\} = q, p + q = 1$. 利用第二章 §10 中引进的依概率收敛的概念, J. 伯努利大数定律可以表述为:

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} p, \quad n \to \infty, \tag{1}$$

其中 $S_n = \xi_1 + \cdots + \xi_n$. (在第四章将证明, 这里实际上也依概率 1 收敛.)

记

$$F_n(x) = \mathbf{P}\left\{\frac{S_n}{n} \leqslant x\right\},$$

$$F(x) = \begin{cases} 1, & 若\ x \geqslant p, \\ 0, & 若\ x < p, \end{cases} \tag{2}$$

其中 $F(x)$ 是退化随机变量 $\xi \equiv p$ 的分布函数. 设 P_n 和 P 是空间 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上对应于分布函数 F_n 和 F 的概率测度.

根据第二章 §10 定理 2, 依概率收敛 $S_n/n \xrightarrow{\mathbf{P}} p$ 必导致按分布收敛 $S_n/n \xrightarrow{d} p$, 而由按分布收敛可见, 对于 $\mathbb{R}$ 上的有界连续函数类 C 中的任何函数 $f = f(x)$, 有

$$\mathbf{E}f\left(\frac{S_n}{n}\right) \to \mathbf{E}f(p). \tag{3}$$

由于

$$\mathbf{E}f\left(\frac{S_n}{n}\right) = \int_{\mathbb{R}} f(x) P_n(\mathrm{d}x), \quad \mathbf{E}f(p) = \int_{\mathbb{R}} f(x) P(\mathrm{d}x),$$

则由 (3) 式, 可以将其写为

$$\int_{\mathbb{R}} f(x) P_n(\mathrm{d}x) \to \int_{\mathbb{R}} f(x) P(\mathrm{d}x), \quad f \in C, \tag{4}$$

或 (根据第二章 §6 的记号) 将其写为

$$\int_{\mathbb{R}} f(x) \mathrm{d}F_n(x) \to \int_{\mathbb{R}} f(x) \mathrm{d}F(x), \quad f \in C. \tag{5}$$

在数学分析中, 收敛性 (4) 称做 (当 $n \to \infty$ 时测度 P_n 向测度 P 的) 弱收敛, 记作 $P_n \xrightarrow{w} P$ (对照定义 2). 自然, 收敛性 (5) 亦称做分布函数 F_n 向 F 的弱收敛, 记作 $F_n \xrightarrow{w} F$.

这样, 在伯努利概型中, 可以断定

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} p \Rightarrow F_n \xrightarrow{w} F, \tag{6}$$

这样, 由 (1) 式亦不难导出, 对于由 (2) 式引进的分布函数, 对于除函数 $F(x)$ 的一个间断点 $x = p$ 之外的所有点 $x \in \mathbb{R}$, 有

$$F_n(x) \to F(x), \quad n \to \infty.$$

这一事实说明, 对于所有点 $x \in \mathbb{R}$, 当 $n \to \infty$ 时弱收敛 $F_n \xrightarrow{w} F$ 并不会导致函数 $F_n(x)$ 向 $F(x)$ 的逐点收敛. 不过, 结果表明, 无论是伯努利概型, 还是任意分布函数的一般情形, 弱收敛等价于下面定义的所谓基本收敛 (见下面定理 2).

定义 1 称定义在数轴上的分布函数序列 $\{F_n\}$, 基本收敛于分布函数 F (记作 $F_n \Rightarrow F$), 如果当 $n \to \infty$ 时,

$$F_n(x) \to F(x), \quad x \in \mathbb{C}(F),$$

其中 $\mathbb{C}(F)$ 是极限函数 F 的连续点的集合.

对于所考虑的伯努利概型, 函数 $F = F(x)$ 是退化的, 而由此不难导出 (参见第二章 §10 练习题 7):

$$(F_n \Rightarrow F) \Rightarrow \left(\frac{S_n}{n} \xrightarrow{\mathbf{P}} p\right)$$

这样, 考虑到下面的定理 2, 有

$$\left(\frac{S_n}{n} \xrightarrow{\mathbf{P}} p\right) \Rightarrow \left(F_n \xrightarrow{w} F\right) \Leftrightarrow (F_n \Rightarrow F) \Rightarrow \left(\frac{S_n}{n} \xrightarrow{\mathbf{P}} p\right), \tag{7}$$

因而大数定律的论断, 可以视为由 (2) 式定义的关于分布函数弱收敛的论点之一.

记

$$\begin{aligned} F_n(x) &= \mathbf{P}\left\{\frac{S_n - np}{\sqrt{npq}} \leqslant x\right\}, \\ F(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u. \end{aligned} \tag{8}$$

根据棣莫弗 - 拉普拉斯定理 (第一章 §6), 对于一切 $x \in \mathbb{R}, F_n(x) \to F(x)$, 从而 $F_n \Rightarrow F$. 由于上面指出的弱收敛 $F_n \xrightarrow{w} F$ 与基本收敛 $F_n \Rightarrow F$ 等价性, 则可以认为棣莫弗 - 拉普拉斯定理, 也是关于 (8) 式定义的分布函数的弱收敛的论断.

这两个例子证实下面定义 2 引进的概率测度弱收敛概念的合理性. 虽然对于数轴的情形, 弱收敛与分布函数的基本收敛等价, 不过作为开始还是倾向于考虑弱收敛: 第一, 因为弱收敛更便于分析; 第二, 因为弱收敛对于比数轴更一般的空间也有意义, 特别, 对于度量空间, 最重要的例子是空间 $\mathbb{R}^n, \mathbb{R}^\infty, C$ 和 D (见第二章 §3).

3. 弱收敛和基本收敛 设 $(E, \mathscr{E}, \rho)$ 是度量空间, $\rho = \rho(x, y)$ 为度量, $\mathscr{E}$ 为由开集生成的 σ- 代数, 而 $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \cdots$ 是空间 $(E, \mathscr{E}, \rho)$ 上的概率测度.

定义 2 称概率测度序列 $\{\mathbf{P}_n\}$ 弱收敛于概率测度 $\mathbf{P}$ (记作 $\mathbf{P}_n \xrightarrow{w} \mathbf{P}$, 其中 w 是英文词 weak convergence (弱收敛) 的字头), 如果对于 $\mathscr{E}$ 上的连续有界函数类 $C(\mathscr{E})$ 中的任意函数 $f = f(x)$, 有

$$\int_E f(x)\mathbf{P}_n(\mathrm{d}x) \to \int_E f(x)\mathbf{P}(\mathrm{d}x). \tag{9}$$

定义 3 称概率测度序列 $\{\mathbf{P}_n\}$ 基本收敛于概率测度 $\mathbf{P}$ (记作 $\mathbf{P}_n \Rightarrow \mathbf{P}$), 如果

$$\mathbf{P}_n(A) \to \mathbf{P}(A). \tag{10}$$

其中 $A \in \mathscr{E}$ 是任意集合, 满足

$$\mathbf{P}(\partial A) = 0, \tag{11}$$

而 ∂A 表示集合 A 的边界:

$$\partial A = [A] \cap [\bar{A}],$$

其中 $[A]$ 表示集合 A 的闭包.

下面的定理表明, 概率测度弱收敛的概念与概率测度基本收敛的概念等价, 以及其他等价的提法.

定理 1 下列各命题等价:

(I) $\mathbf{P}_n \xrightarrow{w} \mathbf{P}$;

(II) 对于闭集 $A, \overline{\lim}\mathbf{P}_n(A) \leqslant \mathbf{P}(A)$;

(III) 对于开集 $A, \underline{\lim}\mathbf{P}_n(A) \geqslant \mathbf{P}(A)$;

(IV) $\mathbf{P}_n \Rightarrow \mathbf{P}$.

证明 (I) $\Rightarrow$ (II). 设 A 是闭集, 记

$$f_A^\varepsilon(x) = \left[1 - \frac{\rho(x, A)}{\varepsilon}\right]^+,$$

其中

$$\rho(x, A) = \inf\{\rho(x, y) : y \in A\}, [x]^+ = \max[0, x].$$

再记

$$A^\varepsilon = \{x : \rho(x, A) < \varepsilon\},$$

并注意到当 $\varepsilon \downarrow 0$ 时 $A^\varepsilon \downarrow A$.

由于函数 $f_A^\varepsilon(x)$ 有界、连续并且满足

$$\mathbf{P}_n(A) = \int_E I_A(x)\mathbf{P}_n(\mathrm{d}x) \leqslant \int_E f_A^\varepsilon(x)\mathbf{P}_n(\mathrm{d}x),$$

则当 $\varepsilon \downarrow 0$ 时

$$\overline{\lim_n}\,\mathbf{P}_n(A) \leqslant \overline{\lim_n} \int_E f_A^\varepsilon(x)\mathbf{P}_n(\mathrm{d}x) = \int_E f_A^\varepsilon(x)\mathbf{P}(\mathrm{d}x) \leqslant \mathbf{P}(A^\varepsilon) \downarrow \mathbf{P}(A),$$

因此 (I) $\Rightarrow$ (II) 得证.

如果由集合转换为其相应的补集, 则蕴涵关系 (II) $\Rightarrow$ (III) 和 (III) $\Rightarrow$ (II) 显然.

(III) $\Rightarrow$ (IV). 设 $A^0 = A \backslash \partial A$ 是集合 A 的内部, 而 $[A]$ 是集合 A 的闭包. 由于 (II) 和 (III) 以及假设 $\mathbf{P}(\partial A) = 0$, 则

$$\overline{\lim_n}\,\mathbf{P}_n(A) \leqslant \overline{\lim_n}\,\mathbf{P}_n([A]) \leqslant \mathbf{P}_n([A]) = \mathbf{P}(A),$$

$$\underline{\lim_n}\,\mathbf{P}_n(A) \geqslant \underline{\lim_n}\,\mathbf{P}_n(A^0) \geqslant \mathbf{P}(A^0) = \mathbf{P}(A),$$

因此对于每个 A, 若 $\mathbf{P}(\partial A)=0$, 则 $\mathbf{P}_n(A)\to\mathbf{P}(A)$.

(Ⅳ) ⇒ (Ⅰ). 设 $f=f(x)$ 是连续有界函数,$|f(x)|<M$. 记

$$D=\{t\in\mathbb{R}:\mathbf{P}\{x:f(x)=t\}\neq 0\};$$

考虑区间 $[-M,M]$ 的分割 $T_k=(t_0,t_1,\cdots,t_k)$:

$$-M=t_0<t_1<\cdots<t_k=M, k\geqslant 1,$$

其中 $t_i\notin D, i=0,1,\cdots,k$. (注意, 由于集合 $f^{-1}\{t\}$ 两两不相交, 而测度 $\mathbf{P}$ 有限, 故集合 D 有限或可数.)

设 $B_i=\{x:t_i\leqslant f(x)<t_{i+1}\}$. 由于函数 $f(x)$ 连续, 可见 $f^{-1}(t_i,t_{i+1})$ 是开集, 则

$$\partial B_i\subseteq f^{-1}\{t_i\}\cup f^{-1}\{t_{i+1}\}.$$

点 $t_i,t_{i+1}\notin D$, 因此 $\mathbf{P}(\partial B_i)=0$, 故由 (IV), 有

$$\sum_{i=0}^{k-1}t_i\mathbf{P}_n(B_i)\to\sum_{i=0}^{k-1}t_i\mathbf{P}(B_i). \tag{12}$$

由于

$$\begin{aligned}\left|\int_E f(x)\mathbf{P}_n(\mathrm{d}x)-\int_E f(x)\mathbf{P}(\mathrm{d}x)\right| &\leqslant \left|\int_E f(x)\mathbf{P}_n(\mathrm{d}x)-\sum_{i=0}^{k-1}t_i\mathbf{P}_n(B_i)\right|\\ &\quad+\left|\sum_{i=0}^{k-1}t_i\mathbf{P}_n(B_i)-\sum_{i=0}^{k-1}t_i\mathbf{P}(B_i)\right|+\left|\sum_{i=0}^{k-1}t_i\mathbf{P}(B_i)-\int_E f(x)\mathbf{P}(\mathrm{d}x)\right|\\ &\leqslant 2\max_{0\leqslant i\leqslant k-1}(t_{i+1}-t_i)+\left|\sum_{i=0}^{k-1}t_i\mathbf{P}_n(B_i)-\sum_{i=0}^{k-1}t_i\mathbf{P}(B_i)\right|,\end{aligned}$$

而由 (12) 式及分割 $T_k(k\geqslant 1)$ 的任意性, 可见

$$\lim_n\int_E f(x)\mathbf{P}_n(\mathrm{d}x)=\int_E f(x)\mathbf{P}(\mathrm{d}x).$$ □

注 1 证明中出现的函数 $f(x)=I_A(x)$ 和 $f_A^{\varepsilon}(x)$ 的蕴涵关系 (Ⅰ) ⇒ (Ⅱ), 相应为上半连续性和一致连续性. 注意到这种情况就不难证明, 定理的每一个条件, 与下列条件之一等价:

(Ⅴ) 对于一切有界一致连续函数 $f(x)$, 有

$$\int_E f(x)\mathbf{P}_n(\mathrm{d}x)\to\int_E f(x)\mathbf{P}(\mathrm{d}x);$$

(Ⅵ) 对于一切满足利普希茨 (R. Lipschitz) 条件 (见 §7 引理 2) 的有界函数 $f(x)$, 有

$$\int_E f(x)\mathbf{P}_n(\mathrm{d}x)\to\int_E f(x)\mathbf{P}(\mathrm{d}x);$$

(Ⅶ) 对于一切有界的上半连续函数 $f(x)(\overline{\lim}_n f(x_n) \leqslant f(x)$, 当 $x_n \to x$ 时), 有

$$\overline{\lim_n} \int_E f(x)\mathbf{P}_n(\mathrm{d}x) \leqslant \int_E f(x)\mathbf{P}(\mathrm{d}x);$$

(Ⅷ) 对于一切有界的下半连续函数 $f(x)(\underline{\lim}_n f(x_n) \geqslant f(x)$, 当 $x_n \to x$ 时), 有

$$\underline{\lim_n} \int_E f(x)\mathbf{P}_n(\mathrm{d}x) \geqslant \int_E f(x)\mathbf{P}(\mathrm{d}x).$$

注 2 定理 1 可以推广到如下情形: 将空间 $(E,\mathscr{E},\rho)$ 上的概率测度 $\mathbf{P}_n$ 和 $\mathbf{P}$, 换成任意有限测度 μ_n 和 μ (未必是概率测度). 对于这样的测度, 类似地引进弱收敛 $\mu_n \xrightarrow{w} \mu$ 的概念和基本收敛 $\mu_n \Rightarrow \mu$ 的概念, 并且像定理 1 一样证明下列条件的等价性:

(I^*) $\mu_n \xrightarrow{w} \mu$;

(II^*) 对于闭集 $A, \overline{\lim_n}\mu_n(A) \leqslant \mu(A)$, 且 $\mu_n(E) \to \mu(E)$;

(III^*) 对于开集 $A, \underline{\lim_n}\mu_n(A) \geqslant \mu(A)$, 且 $\mu_n(E) \to \mu(E)$;

(IV^*) $\mu_n \Rightarrow \mu$.

如果将条件 (V) ~ (Ⅷ) 中的 $\mathbf{P}_n$ 和 $\mathbf{P}$, 相应地换成 μ_n 和 μ, 则 (V^*) ~ (VIII^*) 中的每一个条件, 都等价于 (I^*) ~ (IV^*) 中的任何一个条件.

4. 数轴 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 设 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 是数轴, 其中 $\mathbb{R}$ 是实数轴, 而 $\mathscr{B}(\mathbb{R})$ 是由欧几里得度量 $\rho(x,y)=|x-y|$ 生成的博雷尔 σ- 代数 (见第二章 §2 第 2 小节注 2). 记 $P_n(n \geqslant 1)$ 和 P 是 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的概率测度, 而 $F_n(n \geqslant 1)$ 和 F 是相应 P_n 和 P 的分布函数. 那么, 有下面的定理.

定理 2 下列各条件等价:

(1) $P_n \xrightarrow{w} P$;

(2) $P_n \Rightarrow P$;

(3) $F_n \xrightarrow{w} F$;

(4) $F_n \Rightarrow F$.

证明 由于 (2) ⇔ (1) ⇔ (3), 可见只需证明 (2) ⇔ (4).

如果 $P_n \Rightarrow P$, 则 (特别) 对于所有 $x \in \mathbb{R}(P\{x\}=0)$, 有

$$P_n(-\infty,x] \to P(-\infty,x].$$

这说明 $F_n \Rightarrow F$.

现在设 $F_n \Rightarrow F$. 为证明收敛性 $P_n \Rightarrow P$, (由于定理 1) 只需证明对于任意开集 A, 有 $\underline{\lim}_n P_n(A) \geqslant P(A)$.

假如 A 是开集, 那么存在可数个 (形如 (a,b) 的) 两两不相交开区间 $I_1, I_2, \cdots$, 使 $A=\sum_{k=1}^{\infty} I_k$. 固定 $\varepsilon>0$, 在每一个区间 $I_k=(a_k,b_k)$ 中选择一子区间 $I'_k=(a'_k,b'_k]$,

使 $a_k', b_k' \in \mathbb{C}(F)$, 和 $P(I_k) \leqslant P(I_k') + \varepsilon 2^{-k}$. (由于函数 $F = F(x)$ 的间断点的个数有限或可数, 这样的区间 $I_k(k \geqslant 1)$ 确实存在.) 那么, 根据法图引理, 有

$$\underline{\lim}_n P_n(A) = \underline{\lim}_n \sum_{k=1}^{\infty} P_n(I_k) \geqslant \sum_{k=1}^{\infty} \underline{\lim}_n P_n(I_k) \geqslant \sum_{k=1}^{\infty} \underline{\lim}_n P_n(I_k').$$

然而, 由于

$$P_n(I_k') = F_n(b_k') - F_n(a_k') \to F(b_k') - F(a_k') = P(I_k'),$$

因此

$$\underline{\lim}_n P_n(A) \geqslant \sum_{k=1}^{\infty} P(I_k') \geqslant \sum_{k=1}^{\infty} [P(I_k) - \varepsilon 2^{-k}] = P(A) - \varepsilon,$$

因为 $\varepsilon > 0$ 的任意性, 所以就证明了 "若 A 是开集, 则 $\underline{\lim}_n P_n(A) \geqslant P(A)$". □

5. 定义收敛类 设 $(E, \mathscr{E})$ 是可测空间. 子集系 $\mathscr{K}_0(E) \subseteq \mathscr{E}$ 称做*定义类*, 如果对于空间 $(E, \mathscr{E})$ 上的任意两个概率测度 $\mathbf{P}$ 和 $\mathbf{Q}$, 由等式

$$\mathbf{P}(A) = \mathbf{Q}(A), \quad 对于一切 \quad A \in \mathscr{K}_0(E)$$

可见测度 $\mathbf{P}$ 和 $\mathbf{Q}$ 等同, 即

$$\mathbf{P}(A) = \mathbf{Q}(A), \quad 对于一切 \quad A \in \mathscr{E}.$$

假如 $(E, \mathscr{E}, \rho)$ 是度量空间, 那么子集系 $\mathscr{K}_1(E) \subseteq \mathscr{E}$ 称做*定义收敛类*, 如果是任意测度 $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \cdots$, 则由

$$\mathbf{P}_n(A) \to \mathbf{P}(A), \quad 对于一切 \quad A \in \mathscr{K}_1(E) \quad 且 \quad \mathbf{P}(\partial A) = 0$$

得出

$$\mathbf{P}_n(A) \to \mathbf{P}(A), \quad 对于一切 \quad A \in \mathscr{E}, \quad 且 \quad \mathbf{P}(\partial A) = 0.$$

当 $(E, \mathscr{E}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 时, 作为定义类 $\mathscr{K}_0(\mathbb{R})$, 可以选 "初等" 集合类 $\mathscr{K} = \{(-\infty, x] : x \in \mathbb{R}\}$ (第二章 §3 定理 1). 由定理 2 的 (2) 和 (4) 式的等价性条件可见, $\mathscr{K}$ 类也是定义收敛类.

自然, 对于更一般的空间, 也产生定义收敛类的问题.

对于空间 $\mathbb{R}^n (n \geqslant 2)$, 形如 $(-\infty, x] = (-\infty, x_1] \times \cdots \times (-\infty, x_n], x \in (x_1, \cdots, x_n) \in \mathbb{R}^n$ 的集合的 "初等" 集合类 $\mathscr{K}$ 是定义类 (第二章 §3 定理 2), 也是定义收敛类 (练习题 2).

对于空间 $\mathbb{R}^\infty$, 柱集是 "初等" 集合, 由柱集的概率可以唯一地确定一切博雷尔集的概率 (第二章 §3 定理 3). 结果, 这时柱集类就是定义收敛类 (练习题 3).

似乎可以指望, 对于更一般的空间, 柱集类是定义收敛类. 然而, 一般并非如此.

例如, 考虑具有均匀度量 ρ 的空间 $(C, \mathscr{B}(C), \rho)$ (第二章 §2 第 6 小节). 设 $\mathbf{P}$ 是完全集中在函数 $x(t) \equiv 0 \ (0 \leqslant t \leqslant 1)$ 上的概率测度, 而 $\mathbf{P}_n (n \geqslant 1)$ 是概率测度, 其中

对于每一个 $n \geqslant 1$, 都全部集中在函数 $x_n = x_n(t)$ 上 (见图 35). 不难验证, 对于一切 $\mathbf{P}(\partial A) = 0$ 的柱集 $A, \mathbf{P}_n(A) \to \mathbf{P}(A)$. 但是, 例如, 若取集合

$$A = \left\{\alpha \in C : |\alpha(t)| \leqslant \frac{1}{2}, 0 \leqslant t \leqslant 1\right\} \in \mathscr{B}_0(C),$$

其中 $\mathbf{P}(\partial A) = 0, \mathbf{P}_n(A) = 0, \mathbf{P}(A) = 1$, 因而 $\mathbf{P}_n \nRightarrow \mathbf{P}$.

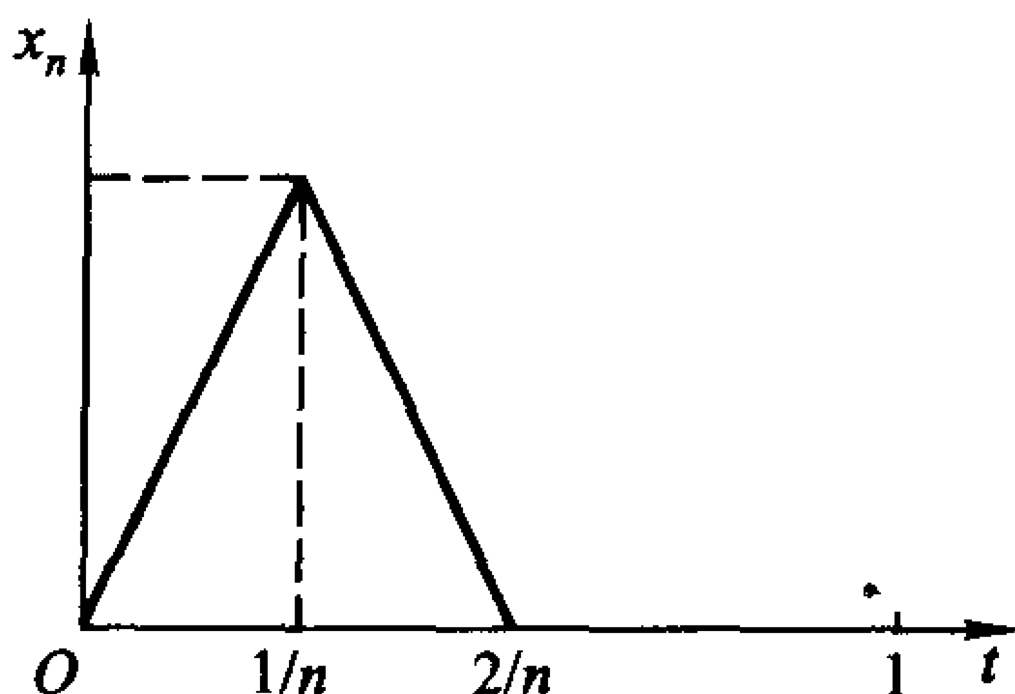


图 35

这样, 对于上面的例子, 柱集类是定义类, 但不是定义收敛类.

6. 练习题

1. 称定义在 $\mathbb{R}^m$ 上的函数 $F = F(x)$在点 $x \in \mathbb{R}^m$ 处连续, 如果对于任何 $\varepsilon > 0$, 存在 $\delta > 0$, 使对于一切满足不等式 $|F(x) - F(y)| < \varepsilon$ 的 $y \in \mathbb{R}^m$, 满足不等式

$$x - \delta e < y < x + \delta e,$$

其中 $e = (1, \cdots, 1) \in \mathbb{R}^m$. 此外, 亦称分布函数序列 $\{F_n\}$基本收敛于分布函数 F (记作 $F_n \Rightarrow F$), 如果在函数 $F = F(x)$ 的所有连续点 $x \in \mathbb{R}^m$, 有 $F_n(x) \to F(x)$.

证明, 对于 $\mathbb{R}^m (m > 1)$, 定理 2 的结论仍然成立. (见定理 1 的注.)

2. 证明, 对于空间 $\mathbb{R}^n$, "初等" 集合类 $\mathscr{K}$ 是定义收敛类.

3. 设 E 是空间 $\mathbb{R}^\infty$, C 或 D 之一, $\mathscr{E}$ 是开集生成的博雷尔集合代数, 而 $\{\mathbf{P}_n\}$ 是定义在 σ- 代数上 $\mathscr{E}$ 的概率测度序列. 称 $\{\mathbf{P}_n\}$ 在有限维分布的意义上基本收敛于概率测度 $\mathbf{P}$ (记作 $\mathbf{P}_n \stackrel{f}{\Rightarrow} \mathbf{P}$), 如果对于一切柱集 $A, \mathbf{P}(\partial A) = 0$, 当 $n \to \infty$ 时 $\mathbf{P}_n(A) \to \mathbf{P}(A)$.

证明, 对于空间为 $\mathbb{R}^\infty$ 的情形, 有

$$(\mathbf{P}_n \stackrel{f}{\Rightarrow} \mathbf{P}) \Leftrightarrow (\mathbf{P}_n \Rightarrow \mathbf{P}).$$

问对于空间 C 和 D, 上述结果是否成立?

4. 设 F 和 G 是数轴上的分布函数, 而

$$L(F, G) = \inf\{h > 0 : F(x-h) - h \leqslant G(x) \leqslant F(x+h) + h\}$$

是 (F 和 G 的) 莱维 (P. P. Lévy) 距离. 证明, "基本收敛" 与按由距离 $L(\cdot,\cdot)$ 决定的 "莱维度量的收敛" 等价:

$$(F_n \Rightarrow F) \Leftrightarrow (L(F_n, F) \to 0).$$

5. 设 $F_n \Rightarrow F$, 且分布函数 F 连续. 证明 $F_n(x)$ 收敛于 $F(x)$, 与当 $n \to \infty$ 时

$$\sup_x |F_n(x) - F(x)| \to 0$$

等价.

6. 证明定理 1 之注 1 提出的命题.

7. 证明定理 1 之注 2 中提出的条件 (I*) ~ (IV*) 等价性的正确性.

8. 证明 $\mathbf{P}_n \xrightarrow{w} \mathbf{P}$, 当且仅当序列 $\{\mathbf{P}_n\}$ 的任何子列 $\{\mathbf{P}_{n'}\}$, 都包含这样的子列 $\{\mathbf{P}_{n''}\}$, 使 $\mathbf{P}_{n''} \xrightarrow{w} \mathbf{P}$.

9. 举出 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上概率测度 $P, P_n(n \geqslant 1)$ 的例满足: $P_n \xrightarrow{w} P$, 但是对于一切博雷尔集合 $B \in \mathscr{B}(\mathbb{R}), P_n(B)$ 可能不收敛于 $P(B)$.

10. 举一分布函数 $F = F(x), F_n = F_n(x), n \geqslant 1$ 的例, 使 $F_n \xrightarrow{w} F$, 但是当 $n \to \infty$ 时, $\sup_x |F_n(x) - F(x)|$ 不收敛于 0.

11. 概率论方面的许多教材, 把定理 2 关于 "分布函数 F_n 收敛于分布函数 F" 的命题 "(4) ⇒ (3)" 与赫利 - 布雷 (E.Helly – J.R.Bray) 的名字相联系. 因此建议证明如下命题:

(a) 赫利 - 布雷引理. 如果 $F_n \Rightarrow F$ (见定义 1), 则

$$\lim_n \int_a^b g(x) \mathrm{d}F_n(x) = \int_a^b g(x) \mathrm{d}F(x),$$

其中 a 和 b 是分布函数 $F = F(x)$ 之连续点的集合中的两个点, 而 $g = g(x)$ 是 $[a, b]$ 上的连续函数.

(b) 赫利 - 布雷定理. 如果 $F_n \Rightarrow F$, 而 $g = g(x)$ 是 $\mathbb{R}$ 上的连续有界函数, 则

$$\lim_n \int_{-\infty}^{\infty} g(x) \mathrm{d}F_n(x) = \int_{-\infty}^{\infty} g(x) \mathrm{d}F(x).$$

12. 设 $F_n \Rightarrow F$, 而对于某个 $b > 0$, 序列

$$\left(\int |x|^b \mathrm{d}F_n(x) \right)_{n \geqslant 1}$$

有界. 证明, 对于 $0 \leqslant a \leqslant b$, 有

$$\lim_n \int |x|^a \mathrm{d}F_n(x) = \int |x|^a \mathrm{d}F(x);$$

对于任意 $k=1,2,\cdots,[b], k\neq b$, 有

$$\lim_n \int x^k \mathrm{d}F_n(x) = \int x^k \mathrm{d}F(x).$$

13. 设 $F_n \Rightarrow F$, 而 $m=\mathrm{med}(F), m_n=\mathrm{med}(F_n)$ 相应为 F 和 F_n 的中位数 (第一章 §4 练习题 5). 假设对于一切 $n\geqslant 1$, 中位数 m 和 m_n 的定义唯一, 证明 $m_n\to m$.

14. 假设函数 F 唯一决定于其各阶矩:

$$a_k=\int_{-\infty}^{\infty} x^k \mathrm{d}F(x),\quad k=1,2,\cdots$$

设分布函数序列 $\{F_n\}_{n\geqslant 1}$ 的矩满足

$$a_{n,k}=\int_{-\infty}^{\infty} x^k \mathrm{d}F_n(x)\to a_k=\int_{-\infty}^{\infty} x^k \mathrm{d}F(x),\quad k=1,2,\cdots$$

证明 $F_n\Rightarrow F$.

15. 证明大数定律的如下的一种形式 —— *辛钦大数定律*: 设 $X_1,X_2,\cdots$ 是具有有限数学期望 $\mathbf{E}X_1=m$ 的、两两独立且同分布的随机变量, 而 $S_n=X_1+\cdots+X_n$, 那么

$$S_n/n \xrightarrow{\mathbf{P}} m.$$

§2. 概率分布族的相对紧性和稠密性

1. 关于分布函数类的列紧性问题　如果给定概率测度序列, 则在考虑它 (弱) 收敛于某个概率测度问题之前, 当然应弄清它一般是否收敛于某个测度, 或者该序列是否至少有一个收敛的子序列.

例如, 对于序列 $\{\mathbf{P}_n\}$, 其中 $\mathbf{P}_{2n}=\mathbf{P}, \mathbf{P}_{2n+1}=\mathbf{Q}$, 而 $\mathbf{P}$ 和 $\mathbf{Q}$ 是不同的概率测度, 那么, $\{\mathbf{P}_n\}$ 显然不收敛, 然而它有两个收敛的子序列 $\{\mathbf{P}_{2n}\}$ 和 $\{\mathbf{P}_{2n+1}\}$.

对于非常简单构造的概率测度 $\mathbf{P}_n(n\geqslant 1)$ 的序列 $\{\mathbf{P}_n\}$, 其中 $\mathbf{P}_n$ 集中在点 $\{n\}$ 上 $(\mathbf{P}_n(\{n\})=1)$, 则 $\{\mathbf{P}_n\}$ 既不收敛, 也不含任何收敛子序列. (因为, 对于任何常数 $a<b, \lim_n \mathbf{P}_n(a,b]=0$, 则极限测度理应恒等于 0, 然而这与当 $n\to\infty$ 时 $1=\mathbf{P}_n(\mathbb{R})\nrightarrow 0$ 矛盾.) 有趣的是, 在这个例子中相应的分布函数序列 $\{F_n\}$:

$$F_n(x)=\begin{cases}1, & 若\ x\geqslant n,\\ 0, & 若\ x<n,\end{cases}$$

显然是收敛的: 对于任何 $x\in\mathbb{R}$,

$$F_n(x)\to G(x)\equiv 0.$$

不过极限函数 $G=G(x)$ (按第二章 §3 定义 1) 并不是分布函数.

这个例子之所以可以借鉴, 是因为它说明分布函数类不是*列紧的*. 这个例子还提示, 为使分布函数序列收敛于也是分布函数的函数, 需要某些预防 “质量在无穷流失” 的条件.

在说明这里遇到的困难的特点之后, 我们给出基本定义.

2. 分布函数类的相对列紧性和完备性 假设所考虑的一切测度定义在度量空间 $(E,\mathscr{E},\rho)$ 上.

定义 1 称概率测度族 $\mathscr{P}=\{\mathbf{P}_\alpha:\alpha\in\mathfrak{A}\}$ 为*相对列紧的*, 如果 $\mathscr{P}$ 中的任何测度序列包含弱收敛于某一概率测度的子序列.

需要强调, 在这一定义中假设极限测度是概率测度, 不过它有可能不属于原概率测度族 $\mathscr{P}$. (正是因为如此, 在该定义中出现了 “相对” 二字.)

验证给定的概率测度族的相对列紧性相当困难. 因此, 希望有进行此检验的简单而方便的准则. 下面的概念有助于实现这一目标.

定义 2 概率测度族 $\mathscr{P}=\{\mathbf{P}_\alpha:\alpha\in\mathfrak{A}\}$ 称为*完备的*, 如果对于每一个 $\varepsilon>0$, 存在紧统 $K\subseteq E$, 使

$$\sup_{\alpha\in\mathfrak{A}}\mathbf{P}_\alpha(E\setminus K)\leqslant\varepsilon. \tag{1}$$

定义 3 定义在 $\mathbb{R}^n(n\geqslant 1)$ 上的分布函数族 $\mathscr{F}=\{F_\alpha:\alpha\in\mathfrak{A}\}$ 称为*相对列紧的 (完备的)*, 如果相应的概率测度族 $\mathscr{P}=\{\mathbf{P}_\alpha:\alpha\in\mathfrak{A}\}$ 是相对列紧的 (完备的), 其中 $\mathbf{P}_\alpha$ 是按 F_α 建立的测度.

3. 概率测度族列紧性的条件 下面的结果, 在整个概率测度的弱收敛问题中有重要的作用.

定理 1 (普罗霍罗夫 [Ю. В. Прохоров] **定理)** *设 $\mathscr{P}=\{\mathbf{P}_\alpha:\alpha\in\mathfrak{A}\}$ 是定义在完全可分度量空间 $(E,\mathscr{E},\rho)$ 上的概率测度族. 测度族 $\mathscr{P}$ 是相对列紧的, 当且仅当它是稠密的.*

证明 我们只对E是数轴的情形证明定理. (这一证明亦可用于任意欧几里得空间 $\mathbb{R}^n(n\geqslant 2)$ ([55],[5]). 然后依次证明对于 $\mathbb{R}^\infty$ 以及对于 σ- 列紧空间定理成立; 最后, 对于一般完备可分度量空间, 通过将每一种情形归结为上一种, 并证明其成立.)

必要性. 假设 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的概率测度族 $\mathscr{P}=\{\mathbf{P}_\alpha:\alpha\in\mathfrak{A}\}$ 相对列紧但不稠密. 那么, 存在这样一个 $\varepsilon>0$, 使对任意紧统 $K\subseteq\mathbb{R}$, 有

$$\sup_\alpha\mathbf{P}_\alpha(\mathbb{R}\backslash K)>\varepsilon,$$

即对于任何区间 $I=(a,b)$, 有

$$\sup_\alpha\mathbf{P}_\alpha(\mathbb{R}\backslash I)>\varepsilon.$$

由此可见, 对于任何区间 $I_n=(-n,n),n\geqslant 1$, 存在这样的测度 $\mathbf{P}_{\alpha_n}$, 使

$$\mathbf{P}_{\alpha_n}(\mathbb{R}\backslash I_n)>\varepsilon.$$

既然测度族 $\mathscr{P}$ 是相对列紧的, 则在序列 $\{\mathbf{P}_{\alpha_n}\}_{n\geqslant 1}$ 中存在子序列 $\{\mathbf{P}_{\alpha_{n_k}}\}$, 使 $\mathbf{P}_{\alpha_{n_k}} \xrightarrow{w} \mathbf{Q}$, 其中 $\mathbf{Q}$ 是某一概率测度.

那么, 由于 §1 中定理 1 的等价条件 (I) 和 (II), 对于任何 $n \geqslant 1$, 有

$$\varlimsup_{k\to\infty} \mathbf{P}_{\alpha_{n_k}}(\mathbb{R}\backslash I_n) \leqslant \mathbf{Q}(\mathbb{R}\backslash I_n). \tag{2}$$

但是 $\mathbf{Q}(\mathbb{R}\backslash I_n) \downarrow 0, n \to \infty$, 而 (2) 式的左侧大于 $\varepsilon > 0$. 这一矛盾说明, 事实上由相对列紧性导致稠密性.

为证明充分性, 我们需要一个 (称为赫利定理的) 一般结果: 关于广义分布函数性质的列紧性 (第二章 §3 第 2 小节). 以 $\mathscr{G} = \{G\}$ 表示满足下列性质的函数 $G = G(x)$ (广义分布函数) 的全体:

1) $G(x)$ 不减;

2) $0 \leqslant G(-\infty), G(+\infty) \leqslant 1$;

3) $G(x)$ 右连续.

显然, $\mathscr{G}$ 包含分布函数类 $\mathscr{F} = \{F\}$, 其中 $F(-\infty) = 0, F(+\infty) = 1$.

定理 2 (赫利定理) *广义分布函数类 $\mathscr{G} = \{G\}$ 是列紧的, 即对于 $\mathscr{G}$ 中的任意函数序列 $\{G\}$, 在 $\mathscr{G}$ 中存在函数 $G(x) \in \mathscr{G}$ 和子序列 $\{n_k\} \subseteq \{n\}$, 使对于属于函数 $G = G(x)$ 之连续点的集合 $\mathbb{C}\{G\}$ 中任何点 x, 有*

$$G_{n_k}(x) \to G(x), k \to \infty.$$

证明 记 $T = \{x_1, x_2, \cdots\}$ 为 $\mathbb{R}$ 上可数且处处稠密的集合. 由于序列 $\{G_n(x_1)\}$ 有界, 故存在子序列 $N_1 = \{n_1^{(1)}, n_2^{(1)}, \cdots\}$, 使当 $i \to \infty$ 时 $G_{n_i^{(1)}}(x_1)$ 收敛于某个数 g_1; 而子序列 N_1 中存在子序列 $N_2 = \{n_1^{(2)}, n_2^{(2)}, \cdots\}$, 使当 $i \to \infty$ 时 $G_{n_i^{(2)}}(x_2)$ 收敛于某个数 g_2, 依此类推.

在集合 $T \subseteq \mathbb{R}$ 上定义函数 $G_T(x)$, 设

$$G_T(x_i) = g_i, x_i \in T,$$

并考虑 "康托尔" 对角序列 $N = \{n_1^{(1)}, n_2^{(2)}, \cdots\}$. 那么, 对于任意 $x_i \in T$, 当 $m \to \infty$ 时, 有

$$G_{n_m^{(m)}}(x_i) \to G_T(x_i).$$

最后, 对于所有 $x \in \mathbb{R}$, 定义函数 $G = G(x)$, 设

$$G(x) = \inf\{G_T(y) : y \in T, y > x\}. \tag{3}$$

可以断定, $G = G(x)$ 就是所求的函数, 且对于 $G(x)$ 的一切连续点 x, 有

$$G_{n_m^{(m)}}(x) \to G(x).$$

事实上, 由于所考虑的一切函数 G_n 都是不减的, 故对于属于集合 T 且满足不等式 $x \leqslant y$ 的一切 x, y, 有

$$G_{n_m^{(m)}}(x) \leqslant G_{n_m^{(m)}}(y).$$

因此, 对于满足上面条件的 x, y, 有 $G_T(x) \leqslant G_T(y)$. 由此和定义 (3), 可见 $G = G(x)$ 是不减函数.

现在证明 $G = G(x)$ 右连续. 设 $x_k \downarrow x$ 和 $d = \lim_k G(x_k)$. 显然 $G(x) \leqslant d$, 只需证明 $G(x) = d$. 假设相反: $G(x) < d$. 那么, 由 (3) 式可见, 存在 $y \in T, x < y$, 使 $G_T(y) < d$. 对于充分大的 $k, x < x_k < y$, 即 $G(x_k) \leqslant G_T(y) < d$, 且 $\lim_k G(x_k) < d$, 而这与 $d = \lim_k G(x_k)$ 矛盾. 于是, 所构造的函数 $G(x)$ 属于 $\mathscr{G}$.

最后, 证明收敛性: 对于任意点 $x^0 \in \mathbb{C}(G)$, 有

$$G_{n_m^{(m)}}(x^0) \to G(x^0).$$

如果 $x^0 < y \in T$, 则

$$\varlimsup_m G_{n_m^{(m)}}(x^0) \leqslant \varlimsup_m G_{n_m^{(m)}}(y) = G_T(y),$$

因此

$$\varlimsup_m G_{n_m^{(m)}}(x^0) \leqslant \inf\{G_T(y) : y \in T, y > x^0\} = G(x^0). \tag{4}$$

另一方面, 设 $x^1 < y < x^0, y \in T$, 则

$$G(x^1) \leqslant G_T(y) = \lim_m G_{n_m^{(m)}}(y) = \varliminf_m G_{n_m^{(m)}}(y) \leqslant \varliminf_m G_{n_m^{(m)}}(x^0).$$

因而, 当 $x^1 \uparrow x^0$ 时, 得

$$G(x^0-) \leqslant \varliminf_m G_{n_m^{(m)}}(x^0). \tag{5}$$

而如果 $G(x^0-) = G(x^0)$, 则由 (4) 和 (5) 式可见

$$G_{n_m^{(m)}}(x^0) \to G(x^0), m \to \infty.$$ □

现在完成定理 1 的证明.

定理 1 的充分性 设概率测度族 $\mathscr{P}$ 完备, 而 $\{\mathbf{P}_n\}$ 是 $\mathscr{P}$ 中的某一概率测度序列, 以 $\{F_n\}$ 表示相应的分布函数序列.

由于赫利定理, 存在 $\{F_n\}$ 的子序列 $\{F_{n_k}\} \subseteq \{F_n\}$ 和广义分布函数 $G(x) \in \mathscr{G}$, 使对于 $x \in \mathbb{C}\{G\}, F_{n_k}(x) \to G(x)$. 现在证明, 由概率测度族 $\mathscr{P}$ 完备的假设, 可见函数 $G = G(x)$ 是 "真正的" 分布函数 $(G(-\infty) = 0, G(+\infty) = 1)$.

取 $\varepsilon > 0$, 设 $I = (a, b]$ 是使

$$\sup_n \mathbf{P}_n(\mathbb{R} \setminus I) < \varepsilon,$$

即满足

$$1-\varepsilon \leqslant \mathbf{P}_n(a,b], n \geqslant 1,$$

的区间. 选择点 $a', b' \in \mathbb{C}(G)$, 使 $a' < a, b' > b$. 那么,

$$1-\varepsilon \leqslant \mathbf{P}_{n_k}(a,b] \leqslant \mathbf{P}_{n_k}(a',b'] = F_{n_k}(b') - F_{n_k}(a') \to G(b') - G(a').$$

由此可见 $G(+\infty) - G(-\infty) = 1$, 而由于 $0 \leqslant G(-\infty) \leqslant G(+\infty) \leqslant 1$, 可见 $G(-\infty) = 0, G(+\infty) = 1$.

于是, 极限函数 $G = G(x)$ 是分布函数, 且 $F_{n_k} \Rightarrow G$. 因此, 连同 §1 的定理 2 就证明了 $\mathbf{P}_{n_k} \xrightarrow{w} \mathbf{Q}$, 其中 $\mathbf{Q}$ 是根据分布函数 G 建立的概率测度. □

4. 练习题

1. 对于空间 $\mathbb{R}^n (n \geqslant 2)$, 证明定理 1 和 2.

2. 设 $\mathbf{P}_\alpha$ 是直线上参数为 m_α 和 $\sigma_\alpha^2, \alpha \in \mathfrak{A}$ 的高斯测度. 证明, 测度族 $\mathscr{P} = \{\mathbf{P}_\alpha; \alpha \in \mathfrak{A}\}$ 是完备的, 当且仅当存在常数 a 和 b, 使

$$|m_\alpha| \leqslant a, \sigma_\alpha^2 \leqslant b, \alpha \in \mathfrak{A}.$$

3. 举例: 定义在 $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ 上的, (1) 稠密的概率测度族 $\mathscr{P} = \{\mathbf{P}_\alpha; \alpha \in \mathfrak{A}\}$, (2) 不稠密的概率测度族 $\mathscr{P} = \{\mathbf{P}_\alpha; \alpha \in \mathfrak{A}\}$.

4. 设 P 是度量空间 $(E, \mathscr{E}, \rho)$ 上的概率测度. 称测度 P 为*稠密的* (对照定义 2), 如果对于任意 $\varepsilon > 0$, 存在紧统 $K \subseteq E$, 使 $P(K) \geqslant 1 - \varepsilon$. 证明下面的结果 (伍拉姆 [Woollam] 定理): *每一个完全可分度量空间上的概率测度 P 是稠密的.*

5. 设 $X = \{X_\alpha, \alpha \in \mathfrak{A}\}$ 是一随机向量族, 且对于某个 $r > 0, \sup_\alpha \|X_\alpha\|^r < \infty$. 证明分布的 $P_\alpha = \mathrm{Law}(X_\alpha)$ 分布族 $\mathscr{P} = \{P_\alpha; \alpha \in \mathfrak{A}\}$ 是稠密的.

§3. 极限定理证明的特征函数法

1. 历史情况概述　在概率论中, 最早极限定理 —— 大数定律, 棣莫弗 - 拉普拉斯定理, 以及伯努利概型中的泊松定理 —— 的证明, 基于求极限前对分布函数 F_n 的*直接分析*, 而 F_n 可以十分简单地由二项概率表示. (在伯努利概型中, 随机变量只有两个可能值, 故本质上可能明显地求出函数 F_n.) 然而, 对于更复杂形式的随机变量, 类似的对函数 F_n 直接分析的方法实际上无法实现.

对于服从任意分布的、独立随机变量之和的极限定理的证明, 首先是切比雪夫实现的.

切比雪夫提出的、现在众所周知的 "切比雪夫不等式", 不但可以用初等方法证明 J. 伯努利大数定律, 而且为独立随机变量之和 $S_n = \xi_1 + \cdots + \xi_n (n \geqslant 1)$ 的大数定律: 对一切 $\varepsilon > 0$, 有

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \frac{\mathbf{E}S_n}{n}\right| \geqslant \varepsilon\right\} \to 0, n \to \infty, \tag{1}$$

建立了相当一般的条件. (见练习题 2.)

此外, 切比雪夫创建 (而马尔可夫完善) 了所谓 “*矩法*”, 用这种方法可以证明棣莫弗 - 拉普拉斯定理的论断:

$$\mathbf{P}\left\{\frac{S_n-\mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}}\leqslant x\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u(=\Phi(x)), \tag{2}$$

并且在如下意义上具有广泛的特点: 它在关于被加随机变量的性质十分一般的假设下成立. 这正是把命题 (2) 称做概率论的*中心极限定理*的根据.

稍后, 李亚普诺夫提出了另一种证明中心极限定理的方法, (源于拉普拉斯的) 这种方法以概率分布的 “特征函数” 的思想为基础. 后来的发展说明, 李亚普诺夫的 “*特征函数方法*”, 在证明最为多种多样的极限定理时, 是十分有效的, 这决定了它的发展和广泛的应用.

这一方法的实质如下.

2. 分布函数与特征函数对应的连续性 我们已经知道 (第二章 §12), 在分布函数与特征函数之间, 存在一一对应关系. 因此, 可以通过特征函数来研究分布函数的性质. 最出色的事实是, 分布函数的弱收敛 $F_n \xrightarrow{w} F$, 与相应特征函数的逐点收敛 $\varphi_n\to\varphi$, 二者等价. 并且, 下面的定理提供了 “关于数轴上分布的弱收敛定理” 证明的基本工具.

定理 1 (连续性定理) *设 $\{F_n\}$ 是分布函数序列: $F_n=F_n(x), x\in\mathbb{R}$, 而 $\{\varphi_n\}$ 是相应的特征函数序列:*

$$\varphi_n(t)=\int_{-\infty}^{\infty}\mathrm{e}^{itx}\mathrm{d}F_n(x), t\in\mathbb{R}.$$

1) 如果 $F_n\xrightarrow{w}F$, 其中 $F=F(x)$ 是某一分布函数, 则 $\varphi_n(t)\to\varphi(t)$, 其中 $\varphi(t)$ 是 $F=F(x)$ 的特征函数.

2) 如果对于每个 $t\in\mathbb{R}$ 存在极限 $\lim_n\varphi_n(t)$, 而函数 $\varphi(t)=\lim_n\varphi_n(t)$ 在 $t=0$ 连续, 则 $\varphi(t)$ 是某一概率分布 $F=F(x)$ 的特征函数, 且

$$F_n\xrightarrow{w}F.$$

证明 将弱收敛的定义分别用于函数 $\operatorname{Re}\mathrm{e}^{itx}$ 和 $\operatorname{Im}\mathrm{e}^{itx}$, 立即可以证明命题 1). 命题 2) 的证明, 要求事先证明几个引理.

引理 1 *设 $\{\mathbf{P}_n\}$ 是稠密概率测度族. 假设序列 $\{\mathbf{P}_n\}$ 的弱收敛子序列 $\{\mathbf{P}_{n'}\}$, 都收敛于同一概率测度 $\mathbf{P}$. 那么, 整个序列 $\{\mathbf{P}_n\}$ 也弱收敛于同一概率测度 $\mathbf{P}$.*

证明 假设结果相反: $\mathbf{P}_n\stackrel{w}{\not\to}\mathbf{P}$. 那么, 存在这样的有界连续函数 $f=f(x)$, 使

$$\int_{\mathbb{R}}f(x)\mathbf{P}_n(\mathrm{d}x)\nrightarrow\int_{\mathbb{R}}f(x)\mathbf{P}(\mathrm{d}x).$$

由此可见, 存在 $\varepsilon > 0$ 和无限数列$\{n'\} \subseteq \{n\}$, 使

$$\left|\int_{\mathbb{R}} f(x)\mathbf{P}_{n'}(\mathrm{d}x) - \int_{\mathbb{R}} f(x)\mathbf{P}(\mathrm{d}x)\right| \geqslant \varepsilon > 0. \tag{3}$$

根据普罗霍罗夫定理 (§2), 由序列 $\{\mathbf{P}_{n'}\}$ 可以选出子序列 $\{\mathbf{P}_{n''}\}$, 使 $\mathbf{P}_{n''} \xrightarrow{w} \mathbf{Q}$, 其中 $\mathbf{Q}$ 是某一概率测度.

根据引理假设 $\mathbf{P} = \mathbf{Q}$, 因而应有

$$\int_{\mathbb{R}} f(x)\mathbf{P}_{n''}(\mathrm{d}x) \to \int_{\mathbb{R}} f(x)\mathbf{P}(\mathrm{d}x),$$

而这与 (3) 式矛盾, 从而引理得证. □

引理 2 设 $\{\mathbf{P}_n\}$ 是 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 稠密概率测度族. 序列 $\{\mathbf{P}_n\}$ 弱收敛于某概率测度, 当且仅当于每个 $t \in \mathbb{R}$ 存在极限 $\lim_n \varphi_n(t)$, 而函数 $\varphi_n(t)$ 测度 $\mathbf{P}_n$ 的特征函数:

$$\varphi_n(t) = \int_{\mathbb{R}} \mathrm{e}^{itx}\mathbf{P}_n(\mathrm{d}x).$$

证明 如果概率测度族 $\{\mathbf{P}_n\}$ 稠密, 则根据普罗霍罗夫定理, 存在数列 $\{\mathbf{P}_{n'}\}$ 和概率测度 $\mathbf{P}$, 使 $\mathbf{P}_{n'} \xrightarrow{w} \mathbf{P}$. 假设结果相反: 整个序列不收敛于 $\mathbf{P}(\mathbf{P}_n \not\xrightarrow{w} \mathbf{P})$. 那么, 根据引理 1, 存在子序列 $\{\mathbf{P}_{n''}\}$ 和概率测度 $\mathbf{Q}$, 使 $\mathbf{P}_{n''} \xrightarrow{w} \mathbf{Q}$, 并且 $\mathbf{P} \neq \mathbf{Q}$.

现在利用: 对于每个 $t \in \mathbb{R}$ 存在极限 $\lim_n \varphi_n(t)$. 那么,

$$\lim_{n'} \int_{\mathbb{R}} \mathrm{e}^{itx}\mathbf{P}_{n'}(\mathrm{d}x) = \lim_{n''} \int_{\mathbb{R}} \mathrm{e}^{itx}\mathbf{P}_{n''}(\mathrm{d}x)$$

因而

$$\int_{\mathbb{R}} \mathrm{e}^{itx}\mathbf{P}(\mathrm{d}x) = \int_{\mathbb{R}} \mathrm{e}^{itx}\mathbf{Q}(\mathrm{d}x), \quad t \in \mathbb{R}.$$

由于特征函数唯一决定分布 (第二章 §12 定理 2), 故 $\mathbf{P} = \mathbf{Q}$ 与假设 $\mathbf{P}_n \not\xrightarrow{w} \mathbf{P}$ 矛盾.

最后, 由弱收敛的定义, 可见引理相反的结论成立. □

下一引理, 根据特征函数在 0 的邻域内的性质, 给出分布函数 "尾部" 的估计.

引理 3 设 $F = F(x)$ 是数轴上的分布函数, 而 $\varphi = \varphi(t)$ 是其特征函数. 那么, 存在常数 $K > 0$, 使对于任何 $a > 0$, 有

$$\int_{|x| \geqslant 1/a} \mathrm{d}F(x) \leqslant \frac{K}{a} \int_0^a [1 - \mathrm{Re}\varphi(t)]\mathrm{d}t. \tag{4}$$

证明 由于

$$\mathrm{Re}\varphi(t) = \int_{-\infty}^{\infty} \cos tx \mathrm{d}F(x),$$

则傅比尼定理, 可见

$$\begin{aligned}
\frac{1}{a}\int_0^a[1-\mathrm{Re}\varphi(t)]\mathrm{d}t &= \frac{1}{a}\int_0^a\left[\int_{-\infty}^{\infty}(1-\cos tx)\mathrm{d}F(x)\right]\mathrm{d}t \\
&= \int_{-\infty}^{\infty}\left[\frac{1}{a}\int_0^a(1-\cos tx)\mathrm{d}t\right]\mathrm{d}F(x) = \int_{-\infty}^{\infty}\left(1-\frac{\sin ax}{ax}\right)\mathrm{d}F(x) \\
&\geqslant \inf_{|y|\geqslant 1}\left(1-\frac{\sin y}{y}\right)\times\int_{|ax|\geqslant 1}\mathrm{d}F(x) = \frac{1}{K}\int_{|x|\geqslant 1/a}\mathrm{d}F(x),
\end{aligned}$$

其中

$$\frac{1}{K} = \inf_{|y|\geqslant 1}\left(1-\frac{\sin y}{y}\right) = 1-\sin 1 \geqslant \frac{1}{7},$$

于是, 当常数 $K=7$ 时 (4) 式肯定成立. □

定理 1 的命题 2) 的证明. 设 $\varphi_n(t)\to\varphi(t), n\to\infty$, 其中 $\varphi(t)$ 在 0 连续. 现在证明, 由此可见, $\{\mathbf{P}_n\}$ 是稠密概率测度族, 其中 $\mathbf{P}_n$ 是对应于分布函数 F_n 的测度.

由于 (4) 式和控制收敛定理, 当 $n\to\infty$ 时, 有

$$\begin{aligned}
\mathbf{P}_n\left\{\mathbb{R}\backslash\left(-\frac{1}{a},\frac{1}{a}\right)\right\} &= \int_{|x|\geqslant\frac{1}{a}}\mathrm{d}F_n(x) \leqslant \frac{K}{a}\int_0^a[1-\mathrm{Re}\varphi_n(t)]\mathrm{d}t \\
&\to \frac{K}{a}\int_0^a[1-\mathrm{Re}\varphi(t)]\mathrm{d}t.
\end{aligned}$$

由于根据定义 $\varphi(t)$ 在 0 连续且 $\varphi(0)=1$, 可见对于任何 $\varepsilon>0$, 存在 $a>0$, 使对于一切 $n\geqslant 1$, 有

$$\mathbf{P}_n\left\{\mathbb{R}\backslash\left(-\frac{1}{a},\frac{1}{a}\right)\right\} \leqslant \varepsilon.$$

从而测度族 $\{\mathbf{P}_n\}$ 稠密, 而引理 2 知存在概率测度 $\mathbf{P}$, 使

$$\mathbf{P}_n \xrightarrow{w} \mathbf{P}.$$

因此

$$\varphi_n(t) = \int_{-\infty}^{\infty}\mathrm{e}^{itx}\mathbf{P}_n(\mathrm{d}x) \to \int_{-\infty}^{\infty}\mathrm{e}^{itx}\mathbf{P}(\mathrm{d}x),$$

即 $\varphi_n(t)\to\varphi(t)$. 于是, $\varphi(t)$ 是概率测度 $\mathbf{P}$ 的特征函数. □

系 *设 $\{F_n\}$ 是分布函数序列, 而 $\{\varphi_n\}$ 是相应的特征函数序列. 此外, 设 F 是分布函数, 而 φ 是其特征函数. 那么, $F_n\xrightarrow{w}F$, 当且仅当对于一切 $t\in\mathbb{R}, \varphi_n(t)\to\varphi(t), n\to\infty$.*

注 设 $\eta,\eta_1,\eta_2,\cdots$ 是随机变量, 且 $F_{\eta_n}\xrightarrow{w}F_\eta$. 那么, 根据第二章 §10 定义 4, 称随机变量 $\eta_1,\eta_2,\cdots$ 依分布收敛于 η, 并记作 $\eta_n\xrightarrow{d}\eta$. 这一记号很直观 ($d$ 是 distribution 的字头), 因此在表述极限定理时, 常认为表达式 $\eta_n\xrightarrow{d}\eta$ 比 $F_{\eta_n}\xrightarrow{w}F_\eta$ 更好.

3. 极限定理证明的特征函数方法 在下一节, 定理 1 将用于不同分布的、独立随机变量中心极限定理的证明. 证明将在所谓林德伯格 (J. W. Linderbege) 条件下进行. 然后证明, 李亚普诺夫条件可以保障林德伯格条件成立. 现在, 我们用特征函数法证明几个简单的极限定理.

定理 2 (辛钦大数定律) 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}|\xi_1| < \infty, \mathbf{E}\xi_1 = m, S_n = \xi_1 + \cdots + \xi_n$, 则 $S_n/n \xrightarrow{\mathbf{P}} m$, 即对于任何 $\varepsilon > 0$, 有

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right\} \to 0, n \to \infty.$$

证明 设 $\varphi(t) = \mathbf{E}\mathrm{e}^{it\xi_1}$ 和 $\varphi_{S_n/n}(t) = \mathbf{E}\mathrm{e}^{it\frac{S_n}{n}}$, 则由随机变量的独立性和第一章 §12 的公式 (6), 有

$$\varphi_{\frac{S_n}{n}}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n.$$

而根据第二章 §12 的 (14) 式, 有

$$\varphi(t) = 1 + itm + o(t), t \to 0$$

因此, 对于任何给定的 $t \in \mathbb{R}$, 有

$$\varphi\left(\frac{t}{n}\right) = 1 + i\frac{t}{n}m + o\left(\frac{1}{n}\right), n \to \infty,$$

从而,

$$\varphi_{\frac{S_n}{n}}(t) = \left[1 + i\frac{t}{n}m + o\left(\frac{1}{n}\right)\right]^n \to \mathrm{e}^{itm}.$$

函数 $\varphi(t) = \mathrm{e}^{itm}$ 在 0 连续, 并且是集中在点 m 的退化概率分布的特征函数. 于是

$$\frac{S_n}{n} \xrightarrow{d} m,$$

即 (见第二章 §10 的练习题 7)

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} m. \qquad \square$$

定理 3 (独立同分布随机变量的中心极限定理) 设 $\xi_1, \xi_2, \cdots$ 是独立同分布 (非退化) 随机变量序列, 且 $\mathbf{E}\xi_1^2 < \infty, S_n = \xi_1 + \cdots + \xi_n$, 则当 $n \to \infty$ 时, 有

$$\mathbf{P}\left\{\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} \leqslant x\right\} \to \Phi(x), x \in \mathbb{R}, \tag{5}$$

其中

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u.$$

证明 设 $\mathbf{E}\xi_1 = m, \mathbf{D}\xi_1 = \sigma^2$, 而

$$\varphi(t) = \mathbf{E}e^{it(\xi_1 - m)}.$$

那么, 如果记

$$\varphi_n(t) = \mathbf{E}e^{it\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}}},$$

则

$$\varphi_n(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.$$

但是, 根据第二章 §12 的 (14) 式, 有

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), t \to 0.$$

因此, 对于任意固定的 t 和 $n \to \infty$, 有

$$\varphi_n(t) = \left[1 - \frac{\sigma^2 t^2}{2n\sigma^2} + o\left(\frac{1}{n}\right)\right]^n \to e^{-\frac{t^2}{2}}.$$

函数 $e^{-t^2/2}$ 是均值为 0 而方差为 1 的正态分布 (记作 $N(0,1)$) 的特征函数, 故由定理 1 可见 (5) 式成立. 根据定理 1 的注, 可以将该结果写成

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} \xrightarrow{d} N(0,1). \tag{6}$$

□

上面两个定理, 涉及“独立同分布 (规范与中心化) 随机变量之和 $S_n = \xi_1 + \cdots + \xi_n$” 的概率的渐近性质. 然而, 为表述泊松定理 (第一章 §6), 不得不考虑更一般的模型, 即所谓随机变量的*系列*概型.

具体地说, 假设对于每个 $n \geqslant 1$, 给定随机变量 $\xi_{n1}, \cdots, \xi_{nn}$. 换句话说, 给定随机变量的 (下) 三角矩阵:

$$\begin{pmatrix} \xi_{11} & & & \\ \xi_{21} & \xi_{22} & & \\ \xi_{31} & \xi_{32} & \xi_{33} & \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

其中每一行 (第一行除外) 的随机变量相互独立. 设 $S_n = \xi_{n1} + \cdots + \xi_{nn}$.

定理 4 (泊松定理) *设对于每个 $n \geqslant 1$, 独立同分布随机变量 $\xi_{n1}, \cdots, \xi_{nn}$, 有*

$$\mathbf{P}\{\xi_{nk} = 1\} = p_{nk}, \mathbf{P}\{\xi_{nk} = 0\} = q_{nk}, 1 \leqslant k \leqslant n,$$

$$p_{nk} + q_{nk} = 1, \max_{1 \leqslant k \leqslant n} p_{nk} \to 0, p_{n1} + \cdots + p_{nn} \to \lambda > 0, n \to \infty.$$

那么,

$$\mathbf{P}\{S_n = m\} \to \frac{\lambda^m}{m!}e^{-\lambda}, m = 0, 1, \cdots \tag{7}$$

证明 因为对于 $1 \leqslant k \leqslant n$,

$$\mathbf{E}e^{it\xi_{nk}} = p_{nk}e^{it} + q_{nk},$$

所以, 当 $n \to \infty$ 时, 有

$$\begin{aligned}\varphi_{S_n}(t) = \mathbf{E}e^{itS_n} &= \prod_{k=1}^{n}(p_{nk}e^{it} + q_{nk}) \\ &= \prod_{k=1}^{n}[1 + p_{nk}(e^{it} - 1)] \to \exp\{\lambda(e^{it} - 1)\}.\end{aligned}$$

函数 $\varphi(t) = \exp\{\lambda(e^{it} - 1)\}$ 是泊松分布的特征函数 (第二章 §12 第 2 小节例 3), 从而 (7) 式得证. 如果以 $\pi(\lambda)$ 表示参数为 λ 的泊松随机变量, 则仿照 (6) 式可以将 (7) 式写成

$$S_n \xrightarrow{d} \pi(\lambda). \qquad \square$$

4. 练习题

1. 对于空间 $\mathbb{R}^n (n \geqslant 2)$ 的情形, 证明定理 1 的命题.

2. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 并且具有有限均值 $\mathbf{E}|\xi_n|$ 和方差 $\mathbf{D}\xi_n$, 而 $\mathbf{D}\xi_n \leqslant K < \infty$, 其中 K 是某一常数. 利用切比雪夫不等式证明大数定律 (1).

3. 在定理 1 的系中, 证明函数族 $\{\varphi_n\}$ 一致连续, 并且在每一个有限区间上有一致收敛 $\varphi_n \to \varphi$.

4. 设 $\varphi_{\xi_n}(t), n \geqslant 1$ 是随机变量 $\xi_n (n \geqslant 1)$ 的特征函数, 证明 $\xi_n \xrightarrow{d} 0$, 当且仅当在点 $t = 0$ 的某邻域内 $\varphi_{\xi_n}(t) \to 1, n \to \infty$.

5. 设 $X_1, X_2, \cdots$ 是 (取值于 $\mathbb{R}^k$ 的) 独立同分布随机向量序列, 且具有 0 均值和 (有限) 协方差矩阵 Γ. 证明

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{d} N(0, \Gamma).$$

(对照定理 3.)

6. 设 $\xi_1, \xi_2, \cdots$ 和 $\eta_1, \eta_2, \cdots$ 是两个随机变量序列, 且对于每一个 n, ξ_n 与 η_n 独立. 假设当 $n \to \infty$时$\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} \eta$, 其中 ξ 与 η 独立. 证明二维随机变量序列 (ξ_n, η_n) 按分布收敛于 (ξ, η).

设 $f = f(x, y)$ 是连续函数, 证明序列 $f = f(\xi_n, \eta_n)$ 按分布收敛于 $f = f(\xi, \eta)$.

7. 举例说明, 定理 1 之命题 2 中 "极限" 特征函数 $\varphi(t) = \lim_n \varphi_n(t)$, 在 0 的连续性条件一般不能减弱. (换句话说, 假设分布 F 的特征函数 $\varphi(t)$ 在 0 不连续, 则可能出现这样的情况: $\varphi_n(t) \to \varphi(t)$, 然而 $F_n \stackrel{w}{\not\to} F$.) 举例说明: 缺少 $\varphi(t)$ 在 0 连续的条件, 可能导致破坏概率分布族 $P_n (n \geqslant 1)$ 的稠密性, 其中 $\varphi_n(t)$ 是 P_n 的特征函数.

§4. 独立随机变量之和的中心极限定理 I. 林德伯格条件

1. 林德伯格条件 在这一节, 将在满足经典的林德伯格条件的传统假设下, 证明独立随机变量 $\xi_1, \xi_2, \cdots, \xi_n (n \geqslant 1)$ 之 (规范与中心化的) 和 S_n 的中心极限定理. 下一节将考虑更一般的情形: 第一, 中心极限定理直接为 "系列形式", 第二, 其证明在满足所谓非经典条件下进行.

定理 1 设 $\xi_1, \xi_2, \cdots$ 是具有有限二阶矩的独立随机变量序列. 假设

$$m_k = \mathbf{E}\xi_k, \sigma_k^2 = \mathbf{D}\xi_k > 0, S_n = \xi_1 + \cdots + \xi_n, D_n^2 = \sum_{k=1}^n \sigma_k^2,$$

而 $F_k = F_k(x)$ 是随机变量 ξ_k 的分布函数.

假设满足 "林德伯格条件": 对于任意 $\varepsilon > 0$,

$$\text{(L)} \qquad \frac{1}{D_n^2} \sum_{k=1}^n \int_{\{x:|x-m_k| \geqslant \varepsilon D_n\}} (x - m_k)^2 \mathrm{d}F_k(x) \to 0, n \to \infty. \tag{1}$$

那么,

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} \xrightarrow{d} N(0,1). \tag{2}$$

证明 不失普遍性可以认为 $m_k = 0 (k \geqslant 1)$. 记

$$\varphi_k(t) = \mathbf{E}\mathrm{e}^{it\xi_k}, T_n = \frac{S_n}{\sqrt{\mathbf{D}S_n}} = \frac{S_n}{D_n}, \varphi_{S_n}(t) = \mathbf{E}\mathrm{e}^{itS_n}, \varphi_{T_n}(t) = \mathbf{E}\mathrm{e}^{itT_n}.$$

那么,

$$\varphi_{T_n}(t) = \mathbf{E}\mathrm{e}^{itT_n} = \mathbf{E}\mathrm{e}^{i\frac{t}{D_n}S_n} = \varphi_{S_n}\left(\frac{t}{D_n}\right) = \prod_{k=1}^n \varphi_k\left(\frac{t}{D_n}\right), \tag{3}$$

而 (由于 §3 定理 1) 为证明 (2) 式, 只需对于任意 $t \in \mathbb{R}$, 证明

$$\varphi_{T_n}(t) \to \mathrm{e}^{-\frac{t^2}{2}}, n \to \infty. \tag{4}$$

取某个 $t \in \mathbb{R}$, 并认为在整个证明过程中 t 是固定的. 由于分解

$$\mathrm{e}^{iy} = 1 + iy + \frac{\theta_1 y^2}{2},$$
$$\mathrm{e}^{iy} = 1 + iy - \frac{y^2}{2} + \frac{\theta_2 |y|^3}{3!},$$

对于每一个实数 y, 其中 $|\theta_1| \leqslant 1, |\theta_2| \leqslant 1 (\theta_1 = \theta_1(y), \theta_2 = \theta_2(y))$, 有

$$\varphi_k(t) = \mathbf{E}\mathrm{e}^{it\xi_k} = \int_{-\infty}^{\infty} \mathrm{e}^{itx} \mathrm{d}F_k(x)$$
$$= \int_{|x| \geqslant \varepsilon D_n} \left(1 + itx + \frac{\theta_1 (tx)^2}{2}\right) \mathrm{d}F_k(x) + \int_{|x| < \varepsilon D_n} \left(1 + itx - \frac{t^2 x^2}{2} + \frac{\theta_2 |tx|^3}{6}\right) \mathrm{d}F_k(x)$$
$$= 1 + \frac{t^2}{2} \int_{|x| \geqslant \varepsilon D_n} \theta_1 x^2 \mathrm{d}F_k(x) - \frac{t^2}{2} \int_{|x| < \varepsilon D_n} x^2 \mathrm{d}F_k(x) + \frac{|t|^3}{6} \int_{|x| < \varepsilon D_n} \theta_2 |x|^3 \mathrm{d}F_k(x),$$

其中用到: 根据假设

$$m_k = \int_{-\infty}^{\infty} x\mathrm{d}F_k(x) = 0.$$

从而

$$\begin{aligned}\varphi_k\left(\frac{t}{D_n}\right) = 1 - \frac{t^2}{2D_n^2}\int_{|x|<\varepsilon D_n} x^2\mathrm{d}F_k(x) + \frac{t^2}{2D_n^2}\int_{|x|\geqslant\varepsilon D_n} \theta_1 x^2\mathrm{d}F_k(x)\\ + \frac{|t|^3}{6D_n^3}\int_{|x|<\varepsilon D_n} \theta_2|x|^3\mathrm{d}F_k(x).\end{aligned} \tag{5}$$

由于

$$\left|\frac{1}{2}\int_{|x|\geqslant\varepsilon D_n} \theta_1 x^2\mathrm{d}F_k(x)\right| \leqslant \frac{1}{2}\int_{|x|\geqslant\varepsilon D_n} x^2\mathrm{d}F_k(x),$$

有

$$\frac{1}{2}\int_{|x|\geqslant\varepsilon D_n} \theta_1 x^2\mathrm{d}F_k(x) = \widetilde{\theta}_1\int_{|x|\geqslant\varepsilon D_n} x^2\mathrm{d}F_k(x), \tag{6}$$

其中 $\widetilde{\theta}_1 = \widetilde{\theta}_1(t,k,n), |\widetilde{\theta}_1| \leqslant 1/2.$

同样,

$$\left|\frac{1}{6}\int_{|x|<\varepsilon D_n} \theta_2|x|^3\mathrm{d}F_k(x)\right| \leqslant \frac{1}{6}\int_{|x|<\varepsilon D_n} \frac{\varepsilon D_n}{|x|}|x|^3\mathrm{d}F_k(x) \leqslant \frac{1}{6}\int_{|x|<\varepsilon D_n} \varepsilon D_n x^2\mathrm{d}F_k(x),$$

即

$$\frac{1}{6}\int_{|x|<\varepsilon D_n} \theta_2|x|^3\mathrm{d}F_k(x) = \widetilde{\theta}_2\int_{|x|<\varepsilon D_n} \varepsilon D_n x^2\mathrm{d}F_k(x), \tag{7}$$

其中 $\widetilde{\theta}_2 = \widetilde{\theta}_2(t,k,n), |\widetilde{\theta}_2| \leqslant 1/6.$

现在, 设

$$A_{kn} = \frac{1}{D_n^2}\int_{|x|<\varepsilon D_n} x^2\mathrm{d}F_k(x), B_{kn} = \frac{1}{D_n^2}\int_{|x|\geqslant\varepsilon D_n} x^2\mathrm{d}F_k(x).$$

那么, 由 (5) ~ (7) 式, 可见

$$\varphi_k\left(\frac{t}{D_n}\right) = 1 - \frac{t^2 A_{kn}}{2} + t^2\widetilde{\theta}_1 B_{kn} + |t|^3\varepsilon\widetilde{\theta}_2 A_{kn} (= 1 + C_{kn}). \tag{8}$$

注意到

$$\sum_{k=1}^{n}(A_{kn} + B_{kn}) = 1, \tag{9}$$

并根据条件 (1), 有

$$\sum_{k=1}^{n} B_{kn} \to 0, n \to \infty. \tag{10}$$

因此, 对于充分大的 n, 有

$$\max_{1\leqslant k\leqslant n}|C_{kn}|\leqslant t^2\varepsilon^2+\varepsilon|t|^3 \tag{11}$$

和

$$\sum_{k=1}^{n}|C_{kn}|\leqslant t^2+\varepsilon|t|^3. \tag{12}$$

现在利用如下事实: 对于复数 z, 若 $|z|\leqslant 1/2$, 则

$$\ln(1+z)=z+\theta|z|^2,$$

其中 $\theta=\theta(z), |\theta|\leqslant 1$, 而 ln 表示对数的主值 $(\ln z=\ln|z|+i\arg z, -\pi<\arg z\leqslant\pi)$. 那么, 对于充分大的 n, 由 (8) 式和 (11) 式, 可见对于充分小的 $\varepsilon>0$,

$$\ln\varphi_k\left(\frac{t}{D_n}\right)=\ln(1+C_{kn})=C_{kn}+\theta_{kn}|C_{kn}|^2,$$

其中 $|\theta_{kn}|\leqslant 1$. 因而, 有 (3) 式可见

$$\frac{t^2}{2}+\ln\varphi_{T_n}(t)=\frac{t^2}{2}+\sum_{k=1}^{n}\ln\varphi_k\left(\frac{t}{D_n}\right)=\frac{t^2}{2}+\sum_{k=1}^{n}C_{kn}+\sum_{k=1}^{n}\theta_{kn}|C_{kn}|^2.$$

由于

$$\frac{t^2}{2}+\sum_{k=1}^{n}C_{kn}=\frac{t^2}{2}\left(1-\sum_{k=1}^{n}A_{kn}\right)+t^2\sum_{k=1}^{n}\widetilde{\theta}_1(t,k,n)B_{kn}+\varepsilon|t|^3\sum_{k=1}^{n}\widetilde{\theta}_2(t,k,n)A_{kn},$$

而由 (9),(10) 两式, 对于任意 $\delta>0$, 存在充分大的 n_0 和 $\varepsilon>0$, 使对一切 $n\geqslant n_0$, 有

$$\left|\frac{t^2}{2}+\sum_{k=1}^{n}C_{kn}\right|\leqslant\frac{\delta}{2}.$$

此外, 由 (11),(12) 两式, 可见

$$\left|\sum_{k=1}^{n}\theta_{kn}|C_{kn}|^2\right|\leqslant\max_{1\leqslant k\leqslant n}|C_{kn}|\times\sum_{k=1}^{n}|C_{kn}|\leqslant(t^2\varepsilon^2+\varepsilon|t|^3)(t^2+\varepsilon|t|^3).$$

因此, 对于充分大的 n 和适当选择的 $\varepsilon>0$, 可以使

$$\left|\sum_{k=1}^{n}\theta_{kn}|C_{kn}|^2\right|\leqslant\frac{\delta}{2},$$

从而

$$\left|\frac{t^2}{2}+\ln\varphi_{T_n}(t)\right|\leqslant\delta.$$

这样, 对于任何实数 t, 有

$$\varphi_{T_n}(t)e^{\frac{t^2}{2}}\to 1, n\to\infty,$$

于是,

$$\varphi_{T_n}(t)\to e^{-\frac{t^2}{2}}, n\to\infty. \qquad \square$$

2. 林德伯格条件的某些特殊形式　考虑满足林德伯格条件 (1) 的某些特殊情形, 从而也就是中心极限定理成立的情形.

a) *李亚普诺夫条件*. 对于某个 $\delta > 0$,

$$\frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} \mathbf{E}|\xi_k - m_k|^{2+\delta} \to 0, n \to \infty. \tag{13}$$

设 $\varepsilon > 0$, 那么

$$\mathbf{E}|\xi_k - m_k|^{2+\delta} = \int_{-\infty}^{\infty} |x - m_k|^{2+\delta} \mathrm{d}F_k(x)$$
$$\geqslant \int_{\{x:|x-m_k|\geqslant \varepsilon D_n\}} |x - m_k|^{2+\delta} \mathrm{d}F_k(x) \geqslant \varepsilon^\delta D_n^\delta \int_{\{x:|x-m_k|\geqslant \varepsilon D_n\}} (x - m_k)^2 \mathrm{d}F_k(x),$$

因此, 当 n 充分大时, 有

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int_{\{x:|x-m_k|\geqslant \varepsilon D_n\}} (x - m_k)^2 \mathrm{d}F_k(x) \leqslant \frac{1}{\varepsilon^\delta} \times \frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} \mathbf{E}|\xi_k - m_k|^{2+\delta}.$$

从而, 李亚普诺夫条件可以保障林德伯格条件成立.

b) *独立同分布情形*. 假设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量, 且存在 $m = \mathbf{E}\xi_1$ 和方差 $0 < \sigma^2 \equiv \mathbf{D}\xi_1 < \infty$, 那么

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int_{\{x:|x-m|\geqslant \varepsilon D_n\}} (x - m)^2 \mathrm{d}F_k(x) = \frac{n}{n\sigma^2} \int_{\{x:|x-m|\geqslant \varepsilon\sigma^2\sqrt{n}\}} (x - m)^2 \mathrm{d}F_1(x) \to 0,$$

因为当 $n \to \infty$ 时, $\{x : |x - m| \geqslant \varepsilon\sigma^2\sqrt{n}\} \downarrow \varnothing$, 而 $\sigma^2 = \mathbf{E}(\xi_1 - m)^2 < \infty$.

于是, 林德伯格条件成立, 从而由 §3 的定理 3 可得已证明的定理 1.

c) *一致有界的情形*. 假设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, 且对于一切 $n \geqslant 1$, 有

$$|\xi_n| \leqslant K < \infty,$$

其中 K 是某一常数, 且当 $n \to \infty$ 时, $D_n \to \infty$.

那么, 根据切比雪夫不等式

$$\int_{\{x:|x-m_k|\geqslant \varepsilon D_n\}} (x - m_k)^2 \mathrm{d}F_k(x) = \mathbf{E}[(\xi_k - m_k)^2 I(|\xi_k - m_k| \geqslant \varepsilon D_n)]$$
$$\leqslant (2K)^2 \mathbf{P}\{|\xi_k - m_k| \geqslant \varepsilon D_n\} \leqslant (2K)^2 \frac{\sigma_k^2}{\varepsilon^2 D_n^2},$$

因此, 当 $n \to \infty$ 时,

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int_{\{x:|x-m_k|\geqslant \varepsilon D_n\}} (x - m_k)^2 \mathrm{d}F_k(x) \leqslant \frac{(2K)^2}{\varepsilon^2 D_n^2} \to 0.$$

于是, 林德伯格条件仍然成立, 从而中心极限定理成立.

3. 系列形式下的林德伯格条件

注 1 设

$$T_n = \frac{S_n - \mathbf{E}S_n}{D_n}, F_{T_n}(x) = \mathbf{P}\{T_n \leqslant x\}.$$

那么, 命题 2 表示当 $n \to \infty$ 时, 对一切 $x \in \mathbb{R}$.

$$F_{T_n}(x) \to \Phi(x).$$

由于 $\Phi(x)$ 是连续函数, 则这里实际上是一致收敛 (§1 练习题 5):

$$\sup_{x\in\mathbb{R}} |F_{T_n}(x) - \Phi(x)| \to 0, n \to \infty. \tag{14}$$

特别, 由此可见

$$\mathbf{P}\{S_n \leqslant x\} - \Phi\left(\frac{x - \mathbf{E}S_n}{D_n}\right) \to 0, n \to \infty.$$

这一结论可以用语言表示为: 对于充分大的 n, 随机变量 S_n 大致服从均值为 $\mathbf{E}S_n$ 和方差为 $D_n^2 = \mathbf{D}S_n$ 的正态分布.

注 2 由于根据注 1, 当 $n \to \infty$ 时关于 x *一致收敛*: $F_{T_n}(x) \to \Phi(x)$, 自然提出关于 (14) 式的收敛速度的问题. 在 $\xi_1, \xi_2, \cdots$ *独立同分布且* $\mathbf{E}|\xi_1|^3 < \infty$ 的情形下, *贝里 – 埃森* (A. C. Berry – C. G. Esseen) *定理* (*不等式*) (§11) 提供了该问题的答案:

$$\sup_x |F_{T_n}(x) - \Phi(x)| \leqslant C\frac{\mathbf{E}|\xi_1 - \mathbf{E}\xi_1|^3}{\sigma^3\sqrt{n}}, \tag{15}$$

其中 C 是任意常数, 其具体值至今未知.(在 [90] 的第五章 §4.3 用不等式估计了该常数: $1/\sqrt{2\pi} \leqslant C \leqslant 0.765\,5$.)

在 §11 将给出 (15) 式的证明.

注 3 现在赋予*林德伯格条件*略有不同 (甚至是更为紧凑) 的形式, 特别方便用于极限定理的 "系列形式".

设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列,

$$m_k = \mathbf{E}\xi_k, \sigma_k^2 = \mathbf{D}\xi_k, D_n^2 = \sum_{k=1}^n \sigma_k^2 > 0(n \geqslant 1), \xi_{nk} = \frac{\xi_k - m_k}{D_n}.$$

考虑到这些记号, 林德伯格条件 (1) 现在具有如下形式:

$$(\mathrm{L}) \qquad \sum_{i=1}^n \mathbf{E}\{\xi_{nk}^2 I(|\xi_{nk}| \geqslant \varepsilon)\} \to 0, n \to \infty. \tag{16}$$

如果 $S_n = \xi_{n1} + \cdots + \xi_{nn}$, 则 $\mathbf{D}S_n = 1$, 而定理 1 可以赋予这样的形式: *若满足条件* (16), 则

$$S_n \xrightarrow{d} N(0,1).$$

在这种形式下, 中心极限定理不需要假设随机变量 ξ_{nk} 具有 $(\xi_k - m_k)/D_n$ 的特殊形式. 具体地说, 用定理 1 的证明方法, 可以证明如下结果:

定理 2 设对于每一个 $n \geqslant 1$,

$$\xi_{n1}, \xi_{n2}, \cdots, \xi_{nn}$$

是独立随机变量序列, 且 $\mathbf{E}\xi_{nk} = 0, \mathbf{D}S_n = 1$, 其中 $S_n = \xi_{n1} + \cdots + \xi_{nn}$.

那么, 林德伯格条件 (16) 是收敛 $S_n \xrightarrow{d} N(0,1)$ 的充分条件.

4. 林德伯格条件的必要性 由于

$$\max_{1 \leqslant k \leqslant n} \mathbf{E}\xi_{nk}^2 \leqslant \varepsilon^2 + \sum_{k=1}^{n} \mathbf{E}[\xi_{nk}^2 I(|\xi_{nk}| \geqslant \varepsilon)],$$

则由林德伯格条件 (16), 可见

$$\max_{1 \leqslant k \leqslant n} \mathbf{E}\xi_{nk}^2 \to 0, n \to \infty. \tag{17}$$

很出色的是, 在条件 (17) 下, 由中心极限定理成立, 就可以自然得出林德伯格条件成立[①].

定理 3 设对于每一个 $n \geqslant 1$,

$$\xi_{n1}, \xi_{n2}, \cdots, \xi_{nn}$$

是独立随机变量序列, 且 $\mathbf{E}\xi_{nk} = 0, \mathbf{D}S_n = 1$, 其中 $S_n = \xi_{n1} + \cdots + \xi_{nn}$. 假设满足条件 (17). 那么, 对于中心极限定理 $S_n \to N(0,1)$ 的成立, 林德伯格条件充分而且必要.

由定理 2 立即得充分性. 为证明其必要性, 需要下面的引理 (对照 §3 的引理 3).

引理 设随机变量 ξ 的分布函数和数字特征为 $F = F(x)$, 而且 $\mathbf{E}\xi = 0, \mathbf{D}\xi = \gamma > 0$. 那么, 对于每一个 $a > 0$,

$$\int_{|x| \geqslant 1/a} x^2 \mathrm{d}F(x) \leqslant \frac{1}{a^2}[\mathrm{Re} f(\sqrt{6}a) - 1 + 3\gamma a^2], \tag{18}$$

其中 $f(t) = \mathbf{E}\mathrm{e}^{it\xi}$ 是 ξ 的特征函数.

证明 有

$$\begin{aligned}
&\mathrm{Re} f(t) - 1 + \frac{1}{2}\gamma t^2 = \frac{1}{2}\gamma t^2 - \int_{-\infty}^{\infty} (1 - \cos tx)\mathrm{d}F(x) \\
&= \frac{1}{2}\gamma t^2 - \int_{|x|<1/a} (1 - \cos tx)\mathrm{d}F(x) - \int_{|x| \geqslant 1/a} (1 - \cos tx)\mathrm{d}F(x) \\
&\geqslant \frac{1}{2}\gamma t^2 - \frac{1}{2}t^2 \int_{|x|<1/a} x^2 \mathrm{d}F(x) - 2a^2 \int_{|x| \geqslant 1/a} x^2 \mathrm{d}F(x) \\
&= \left(\frac{1}{2}t^2 - 2a^2\right) \int_{|x| \geqslant 1/a} x^2 \mathrm{d}F(x).
\end{aligned}$$

[①] 条件 (17) 称做 “均匀小条件”. 在均匀小条件下, 林德伯格条件是独立随机变量序列服从中心极限定理的充分和必要条件. —— 译者

设 $t=\sqrt{6}a$, 得所要证明的 (18) 式. □

证明定理 3 的必要性 设

$$F_{nk}(x)=\mathbf{P}\{\xi_{nk}\leqslant x\}, f_{nk}(t)=\mathbf{E}e^{it\xi_{nk}},$$
$$\mathbf{E}\xi_{nk}=0, \mathbf{D}\xi_{nk}=\gamma_{nk}>0, \tag{19}$$
$$\sum_{k=1}^{n}\gamma_{nk}=1, \max_{1\leqslant k\leqslant n}\gamma_{nk}\to 0, n\to\infty.$$

以 $\ln z$ 表示复数 z 之对数的主值 (即 $\ln z=\ln|z|+i\arg z, -\pi<\arg z\leqslant\pi$).

那么,

$$\ln\prod_{k=1}^{n}f_{nk}(t)=\sum_{k=1}^{n}\ln f_{nk}(t)+2\pi im,$$

其中 $m=m(n,t)$ 是某个整数. 从而,

$$\operatorname{Re}\ln\prod_{k=1}^{n}f_{nk}(t)=\operatorname{Re}\sum_{k=1}^{n}\ln f_{nk}(t). \tag{20}$$

由于

$$\prod_{k=1}^{n}f_{nk}(t)\to e^{-\frac{t^2}{2}},$$

故

$$\left|\prod_{k=1}^{n}f_{nk}(t)\right|\to e^{-\frac{t^2}{2}},$$

因而

$$\operatorname{Re}\ln\prod_{k=1}^{n}f_{nk}(t)=\operatorname{Re}\ln\left|\prod_{k=1}^{n}f_{nk}(t)\right|\to-\frac{1}{2}t^2. \tag{21}$$

对于 $|z|<1$,

$$\ln(1+z)=z-\frac{z^2}{2}+\frac{z^3}{3}-\cdots, \tag{22}$$

而对于 $|z|\leqslant 1/2$,

$$|\ln(1+z)-z|\leqslant|z|^2. \tag{23}$$

由于 (19) 式, 对于每个固定的 t, 所有充分大的 n 和一切 $k=1,2,\cdots,n$, 有

$$|f_{nk}(t)-1|\leqslant\frac{1}{2}\gamma_{nk}t^2\leqslant\frac{1}{2}. \tag{24}$$

因此, 由 (23) 和 (24) 两式, 可见当 $n\to\infty$ 时

$$\left|\sum_{k=1}^{n}\{\ln[1+(f_{nk}(t)-1)]-(f_{nk}(t)-1)\}\right|\leqslant\sum_{k=1}^{n}|f_{nk}(t)-1|^2$$
$$\leqslant\frac{t^2}{4}\max_{1\leqslant k\leqslant n}\gamma_{nk}\times\sum_{k=1}^{n}\gamma_{nk}=\frac{t^2}{4}\max_{1\leqslant k\leqslant n}\gamma_{nk}\to 0,$$

从而

$$\left|\operatorname{Re}\sum_{k=1}^{n}\ln f_{nk}(t)-\operatorname{Re}\sum_{k=1}^{n}[f_{nk}(t)-1]\right|\to 0, n\to\infty. \tag{25}$$

由 (20),(21) 和 (25) 式, 可见

$$\operatorname{Re}\sum_{k=1}^{n}[f_{nk}(t)-1]+\frac{1}{2}t^2=\sum_{k=1}^{n}\left[\operatorname{Re}f_{nk}(t)-1+\frac{1}{2}t^2\gamma_{nk}\right]\to 0, n\to\infty.$$

设 $t=\sqrt{6}a$, 对于每一个 $a>0$, 有

$$\sum_{k=1}^{n}[\operatorname{Re}f_{nk}(\sqrt{6}a)-1+3a^2\gamma_{nk}]\to 0, n\to\infty. \tag{26}$$

最后, 由 (18) 式 (其中设 $a=1/\varepsilon$) 和 (26) 式, 得

$$\begin{aligned}\sum_{k=1}^{n}\mathbf{E}[\xi_{nk}^2 I(|\xi_{nk}|\geqslant\varepsilon)]&=\sum_{k=1}^{n}\int_{|x|\geqslant\varepsilon}x^2\mathrm{d}F_{nk}(x)\\&\leqslant\varepsilon^2\sum_{k=1}^{n}[\operatorname{Re}f_{nk}(\sqrt{6}a)-1+3a^2\gamma_{nk}]\to 0, n\to\infty.\end{aligned}$$

于是, 满足林德伯格条件得证. □

5. 练习题

1. 假设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}\xi_1^2<\infty$, 证明, 当 $n\to\infty$ 时, 有

$$\max\left(\frac{|\xi_1|}{\sqrt{n}},\cdots,\frac{|\xi_n|}{\sqrt{n}}\right)\xrightarrow{d}0.$$

2. 对于伯努利概型, 直接证明, 当 $n\to\infty$ 时,

$$\sup_x|F_{T_n}(x)-\Phi(x)|$$

的数量级为 $1/\sqrt{n}$.

3. 设 $X_1,X_2,\cdots$ 是独立可交换 (顺序的) 随机变量序列 (见第二章 §5 练习题 4), 且 $\mathbf{E}X_i=0,\mathbf{E}X_i^2=1$, 且

$$\operatorname{cov}(X_1,X_2)=\operatorname{cov}(X_1^2,X_2^2). \tag{27}$$

证明, 中心极限定理成立:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\xrightarrow{d}N(0,1). \tag{28}$$

相反, 如果 $\mathbf{E}X_n^2<\infty$ 且 (28) 式成立, 则 (27) 式也成立.

4. *局部中心极限定理.* 设 $X_1, X_2, \cdots$ 是独立同分布随机变量, 而且 $\mathbf{E}X_1 = 0, \mathbf{E}X_1^2 = 1$. 假设其特征函数 $\varphi(t) = \mathbf{E}e^{itX_1}$ 具有性质: 对于某个 $r \geqslant 1$, 有

$$\int_{-\infty}^{\infty} |\varphi(t)|^r \mathrm{d}t < \infty.$$

证明, 随机变量 $S_n/\sqrt{n}$ 的概率分布密度 $f_n(x)$ 存在, 且关于 $x \in \mathbb{R}$ 一致, 有

$$f_n(x) \to (2n)^{-\frac{1}{2}} \mathrm{e}^{-\frac{x^2}{2}}, n \to \infty.$$

问相应的结果对于格点随机变量如何?

5. 设 $X_1, X_2, \cdots$ 是独立同分布随机变量, 且 $\mathbf{E}X_1 = 0, \mathbf{E}X_1^2 = 1$. 假设 $d_1^2, d_2^2, \cdots$ 是非负常数满足: $d_n = o(D_n)$, 其中 $D_n^2 = \sum_{k=1}^n d_k^2$. 证明加权变量 $d_1X_1, d_2X_2, \cdots$ 序列服从中心极限定理:

$$\frac{1}{D_n} \sum_{k=1}^{n} d_k X_k \xrightarrow{d} N(0,1), \quad n \to \infty.$$

6. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量, 且 $\mathbf{E}\xi_1 = 0, \mathbf{E}\xi_1^2 = 1$. 假设 $\{\tau_n\}_{n\geqslant 1}$ 是在集合 $\{1, 2, \cdots\}$ 中取值的随机变量, 且 $\tau_n/n \xrightarrow{\mathbf{P}} c$, 其中 $c > 0$ 是常数. 证明 $(S_n = \xi_1 + \cdots + \xi_n)$

$$\mathrm{Law}(\tau_n^{-1/2} S_{\tau_n}) \to \Phi,$$

即 $\tau_n^{-1/2} S_{\tau_n} \xrightarrow{d} \xi \sim N(0,1)$.(注意, 这里并没有假设序列 $\{\tau_n\}_{n\geqslant 1}$ 以及 $\{\xi_n\}_{n\geqslant 1}$ 独立.)

7. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量, 且 $\mathbf{E}\xi_1 = 0, \mathbf{E}\xi_1^2 = 1$. 证明

$$\mathrm{Law}(n^{-1/2} \max_{1\leqslant m\leqslant n} S_m) \to \mathrm{Law}(|\xi|),$$

其中 $\xi \sim N(0,1)$. 换句话说, 对于 $x > 0$, 证明

$$\mathbf{P}\left\{\frac{1}{\sqrt{n}} \max_{1\leqslant m\leqslant n} S_m \leqslant x\right\} \to \sqrt{\frac{2}{\pi}} \int_0^x \mathrm{e}^{-\frac{y^2}{2}} \mathrm{d}y \left(= \frac{1}{\sqrt{2}} \mathrm{erf}(x)\right).$$

提示: 首先对于对称伯努利随机变量 $\xi_1, \xi_2, \cdots : \mathbf{P}\{\xi_n = \pm 1\} = 1/2$, 证明所提出的命题, 然后证明对于任意满足 $\mathbf{E}\xi_1 = 0, \mathbf{E}\xi_1^2 = 1$ 的独立同分布随机变量 $\xi_1, \xi_2, \cdots$, 有同样的极限分布. (*这里极限分布的形状, 不依赖于* "具有 $\mathbf{E}\xi_n = 0, \mathbf{E}\xi_n^2 = 1$ 的独立同分布随机变量 $\xi_1, \xi_2, \cdots$" 的具体选择, 并称做 "不变原理"; 对照 §7.)

8. 在上题的条件下. 证明

$$\mathbf{P}\{n^{-1/2} \max_{1\leqslant m\leqslant n} |S_m| \leqslant x\} \to H(x), x > 0,$$

其中

$$H(x) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left\{-\frac{(2k+1)^2\pi^2}{8x^2}\right\}.$$

9. 设 $X_1, X_2, \cdots$ 是独立随机变量序列, 且

$$\mathbf{P}\{X_n = \pm n_\alpha\} = \frac{1}{2n^\beta}, \mathbf{P}\{X_n = 0\} = 1 - \frac{1}{n^\beta},$$

其中 $2\alpha > \beta - 1$. 证明林德伯格条件成立当且仅当 $0 \leqslant \beta < 1$.

10. 设 $X_1, X_2, \cdots$ 是独立随机变量序列, $|X_n| \leqslant C_n(\mathbf{P}-\text{a.c.})$, 而 $C_n = o(D_n)$, 其中

$$D_n^2 = \sum_{k=1}^n \mathbf{E}(X_k - \mathbf{E}X_k)^2 \to \infty.$$

证明

$$\frac{S_n - \mathbf{E}S_n}{D_n} \to N(0,1), \text{ 其中 } S_n = X_n + \cdots + X_n.$$

11. 设 $X_1, X_2, \cdots$ 是独立随机变量序列, 且 $\mathbf{E}X_n = 0, \mathbf{E}X_n^2 = \sigma_n^2$. 假设 $\{X_n\}_{n\geqslant 1}$ 服从中心极限定理, 而且对于某个 $k \geqslant 1$, 有

$$\mathbf{E}\left(D_n^{-1/2}\sum_{i=1}^n X_i\right)^k \to \frac{(2k)!}{2^k k!}.$$

证明 k 阶林德伯格条件成立, 即

$$\sum_{j=1}^n \int_{\{|x|>\varepsilon\}} |x|^k \mathrm{d}F_j(x) = o(D_n^k), \varepsilon > 0.$$

(普通的林德伯格条件, 对应于 $k = 2$ 的情形; 见 (1) 式.)

12. 设 $X = X(\lambda)$ 和 $Y = Y(\mu)$ 是参数相应为 $\lambda > 0$ 和 $\mu > 0$ 的独立泊松随机变量. 证明当 $\lambda \to \infty, \mu \to \infty$ 时

$$\frac{[X(\lambda) - \lambda] - [Y(\mu) - \mu]}{\sqrt{X(\lambda) + Y(\mu)}} \sim N(0,1).$$

13. 设对于 $n \geqslant 1, n+1$ 维随机向量 $(X_1^{(n)}, \cdots, X_{n+1}^{(n)})$ 在单位球面上均匀分布. 证明下面的庞加莱 (J. H. Poincaré) 定理成立:

$$\lim_{n\to\infty} \mathbf{P}\{\sqrt{n}X_{n+1}^{(n)} \leqslant x\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u.$$

§5. 独立随机变量之和的中心极限定理 II. 非经典条件

1. 在非经典条件下中心极限定理的例 在 §4 曾经证明, 由林德伯格条件 (16) 引出条件

$$\max_{1\leqslant k\leqslant n} \mathbf{E}\xi_{nk}^2 \to 0$$

成立, 而由此可得所谓极限可忽略性条件 (亦称渐近小性条件), 指的是: 对于任意 $\varepsilon > 0$, 有

$$\max_{1 \leqslant k \leqslant n} \mathbf{P}\{|\xi_{nk}| \geqslant \varepsilon\} \to 0, n \to 0.$$

这样, 可以说, §4 中的定理 1 和 2 在极限可忽略性的条件下, 给出了独立同分布随机变量之和服从中心极限定理的条件. 在各被加项满足极限可忽略性条件的假设下, 极限定理习惯上称做 "经典提法" 的定理. 然而, 不难举出非退化随机变量, 在既不满足林德伯格条件, 也不满足极限可忽略性条件的情况下, 仍然服从中心极限定理的例子.

假设 $\xi_1, \xi_2, \cdots$ 是独立正态分布的随机变量序列, 且 $\mathbf{E}\xi_n = 0, \mathbf{D}\xi_1 = 1, \mathbf{D}\xi_k = 2^{k-2}, k \geqslant 2$. 设 $S_n = \xi_{n1} + \cdots + \xi_{nn}$, 其中

$$\xi_{nk} = \frac{\xi_k}{\sqrt{\sum_{i=1}^{n} \mathbf{D}\xi_i}}.$$

不难验证, 这里无论是林德伯格条件, 还是极限可忽略性条件都不满足. 然而, 由于 S_n 本身服从 $\mathbf{E}S_n = 0, \mathbf{D}S_n = 1$ 的正态分布, 所以中心极限定理显然成立.

在不假设 "经典的" 极限可忽略性条件下, 下面的定理 1 将要给出中心极限定理成立的充分 (和必要) 条件. 在此意义下, 下面提出的条件 (Λ) 正是 §5 的题目所反映的所谓 "非经典" 条件.

2. "非经典" 条件与林德伯格条件的联系 假设对于每个 $n \geqslant 1$, 给定 ("系列形式" 的) 独立随机变量序列:

$$\xi_{n1}, \xi_{n2}, \cdots, \xi_{nn},$$

且 $\mathbf{E}\xi_{nk} = 0, \mathbf{D}\xi_{nk} = \sigma_{nk}^2 > 0, \sum_{k=1}^{n} \sigma_{nk}^2 = 1$. 设 $S_n = \xi_{n1} + \cdots + \xi_{nn}, F_{nk}(x) = \mathbf{P}\{\xi_{nk} \leqslant x\}$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{y^2}{2}} \mathrm{d}y, \Phi_{nk}(x) = \Phi\left(\frac{x}{\sigma_{nk}}\right).$$

定理 1 随机变量和

$$S_n \xrightarrow{d} N(0,1) \tag{1}$$

的充分 (和必要) 条件是, 对于每个 $\varepsilon > 0$,

$$(\Lambda) \qquad \sum_{k=1}^{n} \int_{|x|>\varepsilon} |x||F_{nk}(x) - \Phi_{nk}(x)| \mathrm{d}x \to 0, n \to \infty. \tag{2}$$

下一个定理说明了条件 (Λ) 与经典林德伯格条件

$$(\mathrm{L}) \qquad \sum_{k=1}^{n} \int_{|x|>\varepsilon} x^2 \mathrm{d}F_{nk}(x) \to 0, n \to \infty. \tag{3}$$

之间的联系.

定理 2 1. 林德伯格条件保障条件 (Λ) 成立:

$$(\mathrm{L}) \Rightarrow (\Lambda).$$

2. 如果当 $n \to \infty$ 时,

$$\max_{1 \leqslant k \leqslant n} \mathbf{E}\xi_{nk}^2 \to 0,$$

则条件 (Λ) 保障林德伯格条件 (L) 成立:

$$(\Lambda) \Rightarrow (\mathrm{L}).$$

定理 1 的证明. 证明条件 (Λ) 的必要性相当复杂 ([88],[91],[96]). 我们在这里仅证明条件 (Λ) 的充分性.

设

$$f_{nk}(t) = \mathbf{E}\mathrm{e}^{it\xi_{nk}}, \quad f_n(t) = \mathbf{E}\mathrm{e}^{itS_n},$$
$$\varphi_{nk}(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} \mathrm{d}\Phi_{nk}(x), \quad \varphi(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} \mathrm{d}\Phi(x).$$

由第二章 §12, 可见

$$\varphi_{nk}(t) = \mathrm{e}^{-\frac{t^2\sigma_{nk}^2}{2}}, \quad \varphi(t) = \mathrm{e}^{-\frac{t^2}{2}}.$$

根据 §3 的定理 1, $S_n \xrightarrow{d} N(0,1)$ 当且仅当对于任意实数 $t, f_n(t) \to \varphi(t), n \to \infty$.

有

$$f_n(t) - \varphi(t) = \prod_{k=1}^{n} f_{nk}(t) - \prod_{k=1}^{n} \varphi_{nk}(t).$$

由于 $|f_{nk}(t)| \leqslant 1, |\varphi_{nk}(t)| \leqslant 1$, 可见

$$|f_n(t) - \varphi(t)| = \left| \prod_{k=1}^{n} f_{nk}(t) - \prod_{k=1}^{n} \varphi_{nk}(t) \right| \leqslant \sum_{k=1}^{n} |f_{nk}(t) - \varphi_{nk}(t)|$$
$$= \sum_{k=1}^{n} \left| \int_{-\infty}^{\infty} \mathrm{e}^{itx} \mathrm{d}(F_{nk} - \Phi_{nk}) \right| = \sum_{k=1}^{n} \left| \int_{-\infty}^{\infty} \left(\mathrm{e}^{itx} - itx + \frac{t^2x^2}{2} \right) \mathrm{d}(F_{nk} - \Phi_{nk}) \right|, \quad (4)$$

其中用到

$$\int_{-\infty}^{\infty} x^k \mathrm{d}F_{nk} = \int_{-\infty}^{\infty} x^k \mathrm{d}\Phi_{nk}.$$

对于积分

$$\int_a^b \left(\mathrm{e}^{itx} - itx + \frac{t^2x^2}{2} \right) \mathrm{d}(F_{nk} - \Phi_{nk}),$$

运用分部积分公式 (第二章 §6 定理 11), 得 $x^2[1 - F_{nk}(x) + F_{nk}(-x)] \to 0, x \to \infty, x^2[1 - \Phi_{nk}(x) + \Phi_{nk}(-x)] \to 0, x \to \infty$,

$$\int_{-\infty}^{\infty} \left(\mathrm{e}^{itx} - itx + \frac{t^2x^2}{2} \right) \mathrm{d}(F_{nk} - \Phi_{nk})$$
$$= -it \int_{-\infty}^{\infty} (\mathrm{e}^{itx} - 1 - itx)(F_{nk}(x) - \Phi_{nk}(x)) \mathrm{d}x. \quad (5)$$

由 (4) 和 (5) 式, 得

$$|f_n(t)-\varphi(t)| \leqslant \sum_{k=1}^{n}\left|\int_{-\infty}^{\infty}\left(\mathrm{e}^{itx}-itx+\frac{t^2x^2}{2}\right)\mathrm{d}(F_{nk}-\Phi_{nk})\right|$$

$$=\sum_{k=1}^{n}\left|t\int_{-\infty}^{\infty}(\mathrm{e}^{itx}-1-itx)[F_{nk}(x)-\Phi_{nk}(x)]\mathrm{d}x\right|$$

$$\leqslant \frac{|t|^3}{2}\varepsilon\sum_{k=1}^{n}\int_{|x|\leqslant\varepsilon}|x||F_{nk}(x)-\Phi_{nk}(x)|\mathrm{d}x$$

$$+2t^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}|x||F_{nk}(x)-\Phi_{nk}(x)|\mathrm{d}x$$

$$\leqslant \varepsilon|t|^3\sum_{k=1}^{n}\sigma_{nk}^2+2t^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}|x||F_{nk}(x)-\Phi_{nk}(x)|\mathrm{d}x, \tag{6}$$

其中用到不等式

$$\int_{|x|\leqslant\varepsilon}|x||F_{nk}(x)-\Phi_{nk}(x)|\mathrm{d}x\leqslant 2\sigma_{nk}^2, \tag{7}$$

其正确性利用第二章 §6 的 (71) 式容易证明. 由于 $\varepsilon>0$ 是任意的, 由 (6) 式, 以及条件 (Λ) 可见 $f_n(t)\to\varphi(t), n\to\infty$. □

定理 2 的证明. 1. 根据 §4, 由林德伯格条件 (L), 得 $\max_{1\leqslant k\leqslant n}\sigma_{nk}^2\to 0$. 因此, 注意到 $\sum_{k=1}^n\sigma_{nk}^2=1$, 得

$$\sum_{k=1}^{n}\int_{|x|>\varepsilon}x^2\mathrm{d}\Phi_{nk}(x)\leqslant\int_{|x|>\varepsilon/\sqrt{\max\limits_{1\leqslant k\leqslant n}\sigma_{nk}^2}}x^2\mathrm{d}\Phi(x)\to 0, n\to\infty. \tag{8}$$

由此式连同条件 (L), 可见对于任意 $\varepsilon>0$, 有

$$\sum_{k=1}^{n}\int_{|x|>\varepsilon}x^2\mathrm{d}[F_{nk}(x)+\Phi_{nk}(x)]\to 0, n\to\infty. \tag{9}$$

对于固定的 $\varepsilon>0$, 存在连续可微偶函数 $h=h(x):|h(x)|\leqslant x^2, |h'(x)|\leqslant 4|x|$, 使

$$h(x)=\begin{cases}x^2, & 若\ |x|>2\varepsilon,\\ 0, & 若\ |x|\leqslant 2\varepsilon.\end{cases}$$

对于这样的函数 $h=h(x)$, 由 (9) 式, 可见

$$\sum_{k=1}^{n}\int_{|x|>\varepsilon}h(x)\mathrm{d}[F_{nk}(x)+\Phi_{nk}(x)]\to 0, n\to\infty. \tag{10}$$

利用分部积分法, 由 (10) 式, 得

$$\sum_{k=1}^{n}\int_{x\geqslant\varepsilon}h'(x)\{[1-F_{nk}(x)]+[1-\Phi_{nk}(x)]\}\mathrm{d}x=\sum_{k=1}^{n}\int_{x\geqslant\varepsilon}h(x)\mathrm{d}[F_{nk}+\Phi_{nk}]\to 0,$$

$$\sum_{k=1}^{n}\int_{x\leqslant-\varepsilon}h'(x)[F_{nk}(x)+\Phi_{nk}(x)]\mathrm{d}x=\sum_{k=1}^{n}\int_{x\leqslant-\varepsilon}h(x)\mathrm{d}[F_{nk}+\Phi_{nk}]\to 0.$$

由于当 $|x| \geqslant 2\varepsilon$ 时 $h'(x) = 2x$, 可见

$$\sum_{k=1}^{n} \int_{|x|\geqslant 2\varepsilon} |x|[F_{nk}(x) - \Phi_{nk}(x)]\mathrm{d}x \to 0, n \to \infty.$$

这样, 由于 $\varepsilon > 0$ 的任意性, 可见 (L) $\Rightarrow$ (Λ).

2. 由于 $\max_{1\leqslant k\leqslant n} \sigma_{nk}^2 \to 0$ 和 (8) 式, 对于上面引进的函数 $h = h(x)$, 得

$$\sum_{k=1}^{n} \int_{|x|>\varepsilon} h(x)\mathrm{d}\Phi_{nk}(x) \leqslant \sum_{k=1}^{n} \int_{|x|>\varepsilon} x^2 \mathrm{d}\Phi_{nk}(x) \to 0, n \to \infty. \tag{11}$$

其次, 由分部积分法, 可见

$$\begin{aligned}
&\left|\sum_{k=1}^{n} \int_{|x|\geqslant\varepsilon} h(x)\mathrm{d}[F_{nk} - \Phi_{nk}]\right| \\
\leqslant &\left|\sum_{k=1}^{n} \int_{x\geqslant\varepsilon} h(x)\mathrm{d}[(1 - F_{nk}) - (1 - \Phi_{nk})]\right| + \left|\sum_{k=1}^{n} \int_{x\leqslant-\varepsilon} h(x)\mathrm{d}[F_{nk} - \Phi_{nk}]\right| \\
\leqslant &\sum_{k=1}^{n} \int_{x\geqslant\varepsilon} |h'(x)||(1 - F_{nk}) - (1 - \Phi_{nk})|\mathrm{d}x + \sum_{k=1}^{n} \int_{x\leqslant-\varepsilon} |h'(x)||F_{nk} - \Phi_{nk}|\mathrm{d}x \\
\leqslant &\, 4\sum_{k=1}^{n} \int_{|x|\geqslant\varepsilon} |x||F_{nk}(x) - \Phi_{nk}(x)|\mathrm{d}x.
\end{aligned} \tag{12}$$

由 (11) 和 (12) 式, 可见

$$\sum_{k=1}^{n} \int_{|x|\geqslant 2\varepsilon} x^2 \mathrm{d}F_{nk}(x) \leqslant \sum_{k=1}^{n} \int_{|x|\geqslant\varepsilon} h(x)\mathrm{d}F_{nk}(x) \to 0, n \to \infty.$$

即林德伯格条件 (L) 成立. □

3. 练习题

1. 证明 (5) 式.

2. 验证关系式 (10) 和 (12).

3. 假设 $N = (N_t)_{t\geqslant 0}$ 是在第二章 §9 的第 4 小节中引进的更新过程: $N_t = \sum_{n=1}^{\infty} I(T_n \leqslant t), T_n = \sigma_1 + \cdots + \sigma_n$, 其中 $\sigma_1, \sigma_2, \cdots$ 是独立同分布正随机变量序列. 假设 $\mu = \mathbf{E}\sigma_1 < \infty, 0 < \mathbf{D}\sigma_1 < \infty$, 证明中心极限定理成立:

$$\frac{N_t - t\mu^{-1}}{\sqrt{t\mu^{-3}\mathbf{D}\sigma_1}} \xrightarrow{d} N(0, 1),$$

其中 $N(0, 1)$ 表示均值为 0 而方差为 1 的标准正态分布随机变量.

§6. 无限可分分布和稳定分布

1. 在非经典条件下中心极限定理的例 在 §3 曾经指出, 为表述泊松分布, 不得不考虑*系列形式*: 对于每个 $n \geqslant 1$, 给定随机变量序列 $(\xi_{n,k}), 1 \leqslant k \leqslant n$.

设

$$T_n = \xi_{n,1} + \cdots + \xi_{n,n}, n \geqslant 1. \tag{1}$$

*无限可分分布*概念的产生, 与下面的问题相联系: 如何描述 "可以作为随机变量序列 $\{T_n\}_{n\geqslant 1}$ 的*极限分布*" 的某些分布?

一般讲, 像问题的如此这般的提法, 极限分布可以是任意的. 事实上, 假如 ξ 是某一随机变量, 而 $\xi_{n,1} = \xi, \xi_{n,k} = 0(1 < k \leqslant n)$, 则 $T_n \equiv \xi$, 从而极限分布就是 ξ 的分布, 即分布可以是任意的.

为使极限分布问题更富有内容, 我们在这一节将处处假设, 对于每个 $n \geqslant 1$, 随机变量 $\xi_{n,1}, \cdots, \xi_{n,n}$ 不仅独立而且*同分布*.

注意, 泊松定理 (§3 定理 4) 的情况正是这样. 独立同分布随机变量 $\xi_1, \xi_2, \cdots$ 之和 $S_n = \xi_1 + \cdots + \xi_n (n \geqslant 1)$ 的中心极限定理 (§3 定理 3) 就属于这种情形. 实际上, 如果设

$$\xi_{n,k} = \frac{\xi_k - \mathbf{E}\xi_k}{D_n}, D_n^2 = \mathbf{D}S_n,$$

则

$$T_n = \sum_{k=1}^{n} \xi_{n,k} = \frac{S_n - \mathbf{E}S_n}{D_n}.$$

这样, 正态分布和泊松分布, 可以视为系列形式的极限分布. 如果 $T_n \xrightarrow{d} T$, 则直观上容易理解, 由于 T_n 是独立同分布随机变量之和, 则极限随机变量 T 也应该在同样的意义上是独立同分布随机变量之和. 因此. 引出如下定义.

定义 1 随机变量 T (及其分布函数 F_T 和特征函数 φ_T) 称做*无限可分的*, 如果对于每个 $n \geqslant 1$, 在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上存在独立同分布随机变量 $\eta_1, \cdots, \eta_n$, 使*) $T \stackrel{d}{=} \eta_1 + \cdots + \eta_n$ (或同样地使, $F_T = F_{\eta_1} * \cdots * F_{\eta_n}$ 或 $\varphi_T = (\varphi_{\eta_1})^n$).

注 1 如果随机变量 T 定义的原概率空间是充分 "贫乏的", 可能出现下面的情况: 对于任意 $n \geqslant 1$, 对于某些分布函数 $F^{(n)}$ 及其特征函数 $\varphi^{(n)}$, 分布函数 F_T 及其特征函数 φ_T 可以表现为 $F_T = F^{(n)} * \cdots * F^{(n)}$ (n 次) 和 $\varphi_T = (\varphi^{(n)})^n$, 然而表达式 $T \stackrel{d}{=} \eta_1 + \cdots + \eta_n$ *不可能*. 恰好有 "贫乏" 概率空间的一个例子属于杜布 (J. L. Doob; 见 [103]): 在此概率空间上有一参数为 $\lambda = 1$ 的泊松分布随机变量 (它是无限可分的: $F_T = F^{(n)} * \cdots * F^{(n)}$, 其中 $F^{(n)}$ 是对应于参数为 $\lambda = 1/n$ 的泊松分布的分布函数), 但是不存在参数为 $\lambda = 1/2$ 的泊松分布的随机变量 η_1 和 η_2.

---

*) 记号 "$\xi \stackrel{d}{=} \eta$" 表示随机变量 ξ 和 η 依分布相等 (相合), 即 $F_\xi(x) = F_\eta(x)$, 其中 $F_\xi(x)$ 和 $F_\eta(x)$ 是 ξ 和 η 的分布函数.

考虑到以上的结果, 需要强调, 上面的定义 1 实质上不明显地假设, 原概率空间已经充分 "丰富". 为避开杜布所指出的缺陷 (练习题 11), 原概率空间已经足够 "贫乏".

定理 1 *随机变量 T 可以是和 $T_n=\sum_{k=1}^{n}\xi_{n,k}$ 的依分布收敛的极限, 当且仅当 T 无限可分.*

证明 如果 T 无限可分, 则对于每个 $n\geqslant 1$, 存在独立同分布随机变量 $\xi_{n,1},\cdots,\xi_{n,n}$, 使 $T\overset{d}{=}\xi_{n,1}+\cdots+\xi_{n,n}$, 即 $T\overset{d}{=}T_n, n\geqslant 1$.

相反, 假设 $T_n\xrightarrow{d}T$. 我们现在证明 T 无限可分, 即对于每个 $k\geqslant 1$, 存在独立同分布随机变量 $\eta_1,\cdots,\eta_k$, 使 $T\overset{d}{=}\eta_1+\cdots+\eta_k$.

固定某个 $k\geqslant 1$, 并将变量 $T_{nk}=\sum_{i=1}^{nk}\xi_{nk,i}$ 表示为 $\zeta_n^{(1)}+\cdots+\zeta_n^{(k)}$, 其中

$$\zeta_n^{(1)}=\xi_{nk,1}+\cdots+\xi_{nk,n},\cdots,\zeta_n^{(k)}=\xi_{nk,n(k-1)+1}+\cdots+\xi_{nk,nk}.$$

由于 $T_n\xrightarrow{d}T, n\to\infty$, 可见随机变量 $T_{nk}, n\geqslant 1$ 的分布函数序列相对列紧; 故根据普罗霍罗夫定理知, 随机变量 $T_{nk}, n\geqslant 1$ 的分布函数序列稠密 (§2). 其次, 有

$$[\mathbf{P}\{\zeta_n^{(1)}>z\}]^k=\mathbf{P}\{\zeta_n^{(1)}>z,\cdots,\zeta_n^{(k)}>z\}\leqslant\mathbf{P}\{T_{nk}>kz\}$$

和

$$[\mathbf{P}\{\zeta_n^{(1)}<-z\}]^k=\mathbf{P}\{\zeta_n^{(1)}<-z,\cdots,\zeta_n^{(k)}<-z\}\leqslant\mathbf{P}\{T_{nk}<-kz\}.$$

由这两个不等式和 $T_{nk}, n\geqslant 1$ 的分布族的稠密性, 可见 $\zeta_n^{(1)}, n\geqslant 1$ 的分布族的稠密性. 因此, 存在子序列 $\{n_i\}\subset\{n\}$ 和随机向量 $(\eta_1,\cdots,\eta_k)$, 而不失普遍性可以认为, 定义在原 ("丰富") 概率空间上, 并且

$$(\zeta_{n_j}^{(1)},\cdots,\zeta_{n_j}^{(k)})\xrightarrow{d}(\eta_1,\cdots,\eta_k),$$

或者等价地, 对于任意 $\lambda_1,\cdots,\lambda_k\in\mathbb{R}$, 有

$$\mathbf{E}\mathrm{e}^{i(\lambda_1\zeta_{n_j}^{(1)}+\cdots+\lambda_k\zeta_{n_j}^{(k)})}\to\mathbf{E}\mathrm{e}^{i(\lambda_1\eta_1+\cdots+\lambda_k\eta_k)}.$$

由随机变量 $\zeta_{n_j}^{(1)},\cdots,\zeta_{n_j}^{(k)}$ 独立, 可见

$$\mathbf{E}\mathrm{e}^{i(\lambda_1\zeta_{n_j}^{(1)}+\cdots+\lambda_k\zeta_{n_j}^{(k)})}=\mathbf{E}\mathrm{e}^{i\lambda_1\zeta_{n_j}^{(1)}}\cdots\mathbf{E}\mathrm{e}^{i\lambda_k\zeta_{n_j}^{(k)}}\to\mathbf{E}\mathrm{e}^{i\lambda_1\eta_1}\cdots\mathbf{E}\mathrm{e}^{i\lambda_k\eta_k},$$

即

$$\mathbf{E}\mathrm{e}^{i(\lambda_1\eta_1+\cdots+\lambda_k\eta_k)}=\mathbf{E}\mathrm{e}^{i\lambda_1\eta_1}\cdots\mathbf{E}\mathrm{e}^{i\lambda_k\eta_k},$$

而由第二章 §12 定理 4 知随机变量 $\eta_1,\cdots,\eta_n$ 独立. 同样显然, 它们有相同的分布.

此外,

$$T_{n_jk}=\zeta_{n_j}^{(1)}+\cdots+\zeta_{n_j}^{(k)}\xrightarrow{d}\eta_1+\cdots+\eta_k,$$

同样, 有 $T_{n_j k} \xrightarrow{d} T$. 于是 (练习题 1), 得

$$T \stackrel{d}{=} \eta_1 + \cdots + \eta_k. \qquad \square$$

注 2 如果将这一节开始的条件: 对于每个 $n \geqslant 1$, 随机变量 $\xi_{n,1}, \cdots, \xi_{n,n}$ 同分布, 换成它们满足渐近小 $\max_{1 \leqslant k \leqslant n} \mathbf{P}\{|\xi_{nk}| \geqslant \varepsilon\} = 0$ 的条件, 则定理的结论仍然成立.

2. 随机变量无限可分的充分和必要条件 为验证给定的随机变量 T 是否无限可分, 最简单的是考察其特征函数 $\varphi(t)$ 的形式. 如果对于每个 $n \geqslant 1$, 存在这样的特征函数 $\varphi_n(t)$, 使 $\varphi(t) = [\varphi_n(t)]^n$, 则 T 无限可分.

对于高斯分布的情形,

$$\varphi(t) = \mathrm{e}^{itm}\mathrm{e}^{-\frac{t^2\sigma^2}{2}},$$

若设

$$\varphi_n(t) = \mathrm{e}^{it\frac{m}{n}}\mathrm{e}^{-\frac{t^2\frac{\sigma^2}{n}}{2}},$$

则 $\varphi(t) = [\varphi_n(t)]^n$.

对于泊松分布的情形,

$$\varphi(t) = \mathrm{e}^{\lambda(\mathrm{e}^{it}-1)},$$

若设

$$\varphi_n(t) = \mathrm{e}^{\frac{\lambda}{n}(\mathrm{e}^{it}-1)},$$

则 $\varphi(t) = [\varphi_n(t)]^n$.

如果随机变量 T 服从 Γ 分布, 其密度为

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}\mathrm{e}^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}, & \text{若 } x \geqslant 0, \\ 0, & \text{若 } x < 0, \end{cases}$$

则 (第二章 §12 表 2 – 5) 其特征函数等于

$$\varphi(t) = \frac{1}{(1 - i\beta t)^\alpha}.$$

从而, $\varphi(t) = [\varphi_n(t)]^n$, 其中

$$\varphi_n(t) = \frac{1}{(1 - i\beta t)^{\alpha/n}},$$

说明 T 无限可分.

我们 (不加证明) 引进如下关于 “无限可分分布的特征函数一般形式” 的一般结果.

定理 2 (柯尔莫戈洛夫 – 列维 – 辛钦) 随机变量 T 无限可分, 当且仅当其特征函数具有形式 $\varphi(t)=\exp\psi(t)$, 其中

$$\psi(t)=it\beta-\frac{t^2\sigma^2}{2}+\int_{-\infty}^{\infty}\left(\mathrm{e}^{itx}-1-\frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\mathrm{d}\lambda(x), \tag{2}$$

而 $\beta\in\mathbb{R},\sigma^2\geqslant 0,\lambda$ 是 $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ 上的有限测度, 且 $\lambda(0)=0$.

3. 稳定随机变量 设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量序列, 且 $S_n=\xi_1+\cdots+\xi_n$. 假设存在这样的常数 $b_n,a_n>0$ 和随机变量 T, 使

$$\frac{S_n-b_n}{a_n}\xrightarrow{d}T. \tag{3}$$

试问如何描述 (3) 式中随机变量 T 的全部可能的极限分布?

如果对于独立同分布随机变量 $\xi_1,\xi_2,\cdots$, 有 $0<\sigma^2\equiv\mathbf{D}\xi_1<\infty$, 并且设 $b_n=n\mathbf{E}\xi_1$ 而 $a_n=\sigma\sqrt{n}$, 则由 §4 可见, 随机变量 T 服从标准正态分布 $N(0,1)$.

如果

$$f(x)=\frac{\theta}{\pi(x^2+\theta^2)}$$

是参数为 $\theta>0$ 的柯西分布密度, 而 $\xi_1,\xi_2,\cdots$ 是密度为 $f(x)$ 的独立随机变量, 则其特征函数为 $\varphi_{\xi_1}(t)=\mathrm{e}^{-\theta|t|}$, 而

$$\varphi_{S_n/n}(t)=\left(\mathrm{e}^{-\frac{\theta}{n}|t|}\right)^n=\mathrm{e}^{-\theta|t|},$$

因此随机变量 S_n/n 服从柯西分布 (分布参数仍然是 θ).

这样, 作为极限分布, 除了正态分布之外, 还可能有其他分布 (例如柯西分布). 设

$$\frac{S_n-b_n}{a_n}=\sum_{k=1}^{n}\xi_{n,k}=(T_n),\ \text{其中}\ \xi_{n,k}=\frac{\xi_k}{a_n}-\frac{b_n}{na_n},\quad 1\leqslant k\leqslant n.$$

于是, 对于 T, 一切可以意料到的, 可以作为极限分布 (3) 式出现的分布, (根据定理 1) 一定是无限可分分布. 不过所考虑的随机变量 $T_n=(S_n-b_n)/a_n$ 的特点, 为这里可能出现的极限分布的结构, 可能提供了得到补充信息.

为此 (并考虑到注 1), 我们引进如下概念.

定义 2 随机变量 T (及其分布函数 $F(x)$ 和特征函数 $\varphi(t)$) 称做*稳定的*, 如果对于任何 $n\geqslant 1$, 存在这样的常数 $a_n>0,b_n$, 和与 T 同分布的独立随机变量 $\xi_1,\cdots,\xi_n$, 使

$$a_nT+b_n\overset{d}{=}\xi_1+\cdots+\xi_n, \tag{4}$$

或同样地

$$F\left(\frac{x-b_n}{a_n}\right)=F*\cdots*F(x)\quad\text{或}\quad|\varphi(t)|^n=|\varphi(a_nt)|\mathrm{e}^{ib_nt}. \tag{5}$$

定理 3 随机变量 T 可以是随机变量

$$\frac{S_n - b_n}{a_n}, a_n > 0,$$

按分布收敛的极限, 当且仅当 T 是稳定的.

证明 如果 T 是稳定的, 则根据 (4) 式

$$T \stackrel{d}{=} \frac{S_n - b_n}{a_n}, \text{ 从而 } \frac{S_n - b_n}{a_n} \xrightarrow{d} T,$$

其中 $S_n = \xi_1 + \cdots + \xi_n$.

相反, 假设 $\xi_1, \xi_2, \cdots$ 是独立同分布的随机变量序列, $S_n = \xi_1 + \cdots + \xi_n$, 且

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} T, a_n > 0.$$

现在证明 T 是稳定随机变量.

假如 T 是退化随机变量, 则它显然是稳定的. 因此, 我们将假设 T 是非退化随机变量.

固定 $k \geqslant 1$, 记

$$S_n^{(1)} = \xi_1 + \cdots + \xi_n, \cdots, S_n^{(k)} = \xi_{n(k-1)+1} + \cdots + \xi_{kn},$$
$$T_n^{(1)} = \frac{S_n^{(1)} - b_n}{a_n}, \cdots, T_n^{(k)} = \frac{S_n^{(k)} - b_n}{a_n}.$$

显然, 所有随机变量 $T_n^{(1)}, \cdots, T_n^{(k)}$ 都同分布, 且

$$T_n^{(i)} \xrightarrow{d} T, n \to \infty, i = 1, \cdots, k.$$

记

$$U_n^{(k)} = T_n^{(1)} + \cdots + T_n^{(k)}.$$

那么, 同定理 1 的证明一样, 可得

$$U_n^{(k)} \xrightarrow{d} T^{(1)} + \cdots + T^{(k)},$$

其中 $T^{(i)} (1 \leqslant i \leqslant k)$ 独立, 且 $T^{(1)} \stackrel{d}{=} \cdots \stackrel{d}{=} T^{(k)} \stackrel{d}{=} T$.

另一方面,

$$\begin{aligned} U_n^{(k)} &= \frac{\xi_1 + \cdots + \xi_{kn} - kb_n}{a_n} \\ &= \frac{a_{kn}}{a_n}\left(\frac{\xi_1 + \cdots + \xi_{kn} - b_{kn}}{a_{kn}}\right) + \frac{b_{kn} - kb_n}{a_n} \\ &= \alpha_n^{(k)} V_{kn} + \beta_n^{(k)}, \end{aligned} \tag{6}$$

其中

$$\alpha_n^{(k)} = \frac{a_{kn}}{a_n}, \beta_n^{(k)} = \frac{b_{kn} - kb_n}{a_n},$$

而

$$V_{kn} = \frac{\xi_1 + \cdots + \xi_{kn} - b_{kn}}{a_{kn}}$$

由 (6) 式可见

$$V_{kn} = \frac{U_n^{(k)} - \beta_n^{(k)}}{\alpha_n^{(k)}},$$

其中 $V_{kn} \xrightarrow{d} T, U_n^{(k)} \xrightarrow{d} T^{(1)} + \cdots + T^{(k)}, n \to \infty$.

由下面的引理知, 存在常数 $\alpha^{(k)} > 0$ 和 $\beta^{(k)}$, 使 $\alpha_n^{(k)} \to \alpha^{(k)}, \beta_n^{(k)} \to \beta^{(k)}, n \to \infty$, 而

$$T \stackrel{d}{=} \frac{T^{(1)} + \cdots + T^{(k)} - \beta^{(k)}}{\alpha^{(k)}},$$

从而, 随机变量 T 的稳定性得证. □

引理 *设 $\xi_n \xrightarrow{d} \xi$ 且存在常数 $a_n > 0$ 和 b_n, 使*

$$a_n\xi_n + b_n \xrightarrow{d} \widetilde{\xi},$$

而且 ξ 和 $\widetilde{\xi}$ 都是非退化随机变量. 那么, 存在常数 $a > 0$ 和 b, 使 $\lim a_n = a, \lim b_n = b$ 且

$$\widetilde{\xi} \stackrel{d}{=} a\xi + b.$$

证明 设 φ_n, φ 和 $\widetilde{\varphi}$ 相应为 ξ_n, ξ 和 $\widetilde{\xi}$ 的特征函数. 那么, $a_n\xi_n + b_n$ 的特征函数 $\varphi_{a_n\xi_n+b_n}(t)$ 等于 $\mathrm{e}^{itb_n}\varphi_n(a_nt)$. 根据定理 1 的系, 以及 §3 的练习题 3, 有

$$\mathrm{e}^{itb_n}\varphi_n(a_nt) \to \widetilde{\varphi}(t), \tag{7}$$

$$\varphi_n(t) \to \varphi(t), \tag{8}$$

其中收敛性为在 t 的每一变化区间上一致收敛.

设 $\{n_i\}$ 是使 $a_{n_i} \to a$ 的 $\{n\}$ 的子序列. 我们首先证明 $a < \infty$. 假设 $a = \infty$. 由 (7) 式可见, 对于任意 $c > 0$,

$$\sup_{|t| \leqslant c} ||\varphi_n(a_nt)| - |\widetilde{\varphi}(t)|| \to 0, n \to \infty.$$

将 t 换成 $t_{n_i} = t_0/a_{n_i}$. 那么, 由于 $a_{n_i} \to a = \infty$, 可见

$$\left|\varphi_{n_i}\left(a_{n_i}\frac{t_0}{a_{n_i}}\right)\right| - \left|\widetilde{\varphi}\left(\frac{t_0}{a_{n_i}}\right)\right| \to 0,$$

因而

$$|\varphi_{n_i}(t_0)| \to |\widetilde{\varphi}(0)| = 1.$$

然而 $|\varphi_{n_i}(t_0)| \to |\varphi(t_0)|$. 因为对于任意 $t_0 \in \mathbb{R}, |\varphi(t_0)| = 1$, 故根据第二章 §12 定理 5, 随机变量应该是退化的, 而这与引理的假设矛盾.

这样 $a < \infty$. 现在假设存在两个子序列 $\{n_i\}$ 和 $\{n_i'\}$, 使 $a_{n_i} \to a, a_{n_i'} \to a'$, 其中 $a \neq a'$ (为确定性, 设 $0 \leqslant a' < a$). 那么, 由 (7) 和 (8) 式, 可见,

$$|\varphi_{n_i}(a_{n_i}t)| \to |\varphi(at)|, |\varphi_{n_i}(a_{n_i}t)| \to |\widetilde{\varphi}(t)|$$

和

$$|\varphi_{n_i'}(a_{n_i'}t)| \to |\varphi(a't)|, |\varphi_{n_i'}(a_{n_i'}t)| \to |\widetilde{\varphi}(t)|.$$

故

$$|\varphi(at)| = |\varphi(a't)|,$$

从而, 对于任意 $t \in \mathbb{R}$, 有

$$|\varphi(t)| = \left|\varphi\left(\frac{a'}{a}t\right)\right| = \cdots = \left|\varphi\left(\left(\frac{a'}{a}\right)^n t\right)\right| \to 1, n \to \infty.$$

因此 $|\varphi(t)| \equiv 1$, 而根据第二章 §12 定理 5, 由此可见 ξ 是退化随机变量. 所产生的矛盾说明 $a = a'$, 而这表明存在有限极限 $\lim a_n = a$, 并且 $a \geqslant 0$.

现在证明, 存在极限 $\lim b_n = b$, 并且 $a > 0$. 由于 (8) 式在每个有限区间上一致成立, 故

$$\varphi_n(a_n t) \to \varphi(at).$$

因而, 由于 (7) 式可见, 对于所有使 $\varphi(at) \neq 0$ 的 t, 存在极限 $\lim_n \mathrm{e}^{itb_n}$. 设 $\delta > 0$, 使得对于一切 $|t| < \delta$, 有 $\varphi(at) \neq 0$. 那么, 对于这样的 t, 存在极限 $\lim \mathrm{e}^{itb_n}$. 由此可见 (练习题 9) $\overline{\lim}|b_n| < \infty$.

假设存在两个子序列 $\{n_i\}$ 和 $\{n_i'\}$, 使 $\lim b_{n_i} = b$ 和 $\lim b_{n_i'} = b'$. 那么, 对于 $|t| < \delta$, 有

$$\mathrm{e}^{itb} = \mathrm{e}^{itb'},$$

从而 $b = b'$. 于是, 存在有限极限 $\lim b_n = b$, 而根据 (7) 式,

$$\widetilde{\varphi}(t) = \mathrm{e}^{itb}\varphi(at),$$

说明 $\widetilde{\xi} \stackrel{d}{=} a\xi + b$. 由于 $\widetilde{\xi}$ 非退化, 可见 $a > 0$. □

4. 稳定分布特征函数的一般形式 我们现在 (不加证明) 给出稳定分布特征函数的一般形式.

定理 4 (列维 – 辛钦表现) 随机变量 T 是稳定的, 当且仅当其特征函数 $\varphi(t)$ 具有 $\varphi(t) = \exp\psi(t)$ 的形式, 其中

$$\psi(t) = it\beta - d|t|^\alpha \left(1 + i\theta\frac{t}{|t|}G(t, \alpha)\right), \tag{9}$$

其中 $0<\alpha<2, \beta\in\mathbb{R}, d\geqslant 0, |\theta|\leqslant 1$, 且当 $t=0$ 时 $t/|t|=0$, 而

$$G(t,\alpha)=\begin{cases}\tan\dfrac{\pi}{2}\alpha, & 若\ \alpha\neq 1,\\ \dfrac{2}{\pi}\ln|t|, & 若\ \alpha=1.\end{cases} \tag{10}$$

我们指出, 对称稳定分布的特征函数的结构特别简单:

$$\varphi(t)=\mathrm{e}^{-d|t|^{\alpha}}, \tag{11}$$

其中 $0<\alpha\leqslant 2, d\geqslant 0$.

5. 练习题

1. 证明, 如果 $\xi_n\xrightarrow{d}\xi$ 和 $\eta_n\xrightarrow{d}\eta$, 则 $\xi\overset{d}{=}\eta$.

2. 证明, 如果 φ_1 和 φ_2 是两个无限可分特征函数, 则 $\varphi_1\times\varphi_2$ 也是无限可分特征函数.

3. 设 φ_n 是两个无限可分特征函数, 且对于每个 $t\in\mathbb{R}, \varphi_n(t)\to\varphi(t)$, 其中 $\varphi(t)$ 是一特征函数. 证明 $\varphi(t)$ 无限可分.

4. 证明, 无限可分分布的特征函数不等于 0.

5. 试举一例, 随机变量是无限可分的, 但并不是稳定的.

6. 设 ξ 是稳定随机变量, 证明, 对于一切 $r\in(0,\alpha), 0<\alpha<2$, 数学期望 $\mathbf{E}|\xi|^r<\infty$.

7. 证明, 如果 ξ 是参数为 $0<\alpha\leqslant 1$ 的稳定随机变量, 则其特征函数 $\varphi(t)$ 在 $t=0$ 不可微.

8. 直接证明, 函数 $\varphi(t)=\mathrm{e}^{-d|t|^{\alpha}}$ 是特征函数, 其中 $0<\alpha\leqslant 2, d\geqslant 0$.

9. 设 $\{b_n\}_{n\geqslant 1}$ 是一数列, 满足条件: 对于一切 $|t|<\delta, \delta>0$, 极限 $\lim_n \mathrm{e}^{itb_n}$ 存在. 证明 $\overline{\lim}_n|b_n|<\infty$.

10. 证明二项分布和均匀分布, 都不是无限可分分布.

11. 假设分布函数 F 及其特征函数 φ 可以表示为: $F^{(n)}*\cdots*F^{(n)}$(n 次), $\varphi=[\varphi^{(n)}]^n$, 其中 $F^{(n)}$ 是某一分布函数, 而 $\varphi^{(n)}(n\geqslant 1)$ 是其特征函数. 证明存在 (充分 "丰富" 的) 概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 和定义在它上面的随机变量 T 和 $\{\eta_k^n\}_{k\leqslant n}.(n\geqslant 1)$, 满足

$$T\overset{d}{=}\eta_1^{(n)}+\cdots+\eta_n^{(n)}(n\geqslant 1)$$

(T 服从分布 F, 而 $\eta_1,\cdots,\eta_n$ 独立同服从分布 $F^{(n)}$).

12. 举一随机变量的例子, 使之本身不是无限可分的, 然而其特征函数仍然不等于 0.

§7. 弱收敛的"可度量性"

1. 关于弱收敛的可度量性 设 $(E,\mathscr{E},\rho)$ 是度量空间, 而 $\mathscr{P}(E)=\{P\}$ 是 $(E,\mathscr{E})$ 上的概率测度族. 自然地提出问题: 在 §1 中所考虑的弱收敛 $P_n \xrightarrow{w} P$, 是否可以"度量", 即是否可以在 $\mathscr{P}(E)$ 中的两个测度 $\widetilde{P}$ 和 P 之间, 引进这样的距离 $\delta(P,\widetilde{P})$, 使收敛性 $\delta(P_n,P)\to 0$ 与收敛性 $P_n \xrightarrow{w} P$ 等价.

鉴于上面提出的问题, 有意地指出, 随机变量序列依概率收敛 $\xi_n \xrightarrow{\mathbf{P}} \xi$, 可以度量化, 例如, 借助距离

$$d_{\mathbf{P}}(\xi,\eta)=\inf\{\varepsilon>0:\mathbf{P}\{|\xi-\eta|\geqslant\varepsilon\}\leqslant\varepsilon\},$$

或者, 借助距离

$$d(\xi,\eta)=\mathbf{E}\min(1,|\xi-\eta|),\quad d(\xi,\eta)=\mathbf{E}\frac{|\xi-\eta|}{1+|\xi-\eta|}.$$

(更一般地, 可以设 $d(\xi,\eta)=\mathbf{E}g(|\xi-\eta|)$, 其中函数 $g(x),x\geqslant 0$, 可以取在 0 连续的任意博雷尔非负递增函数: 当 $x>0$ 时 $g(x)>0, g(0)=0$, 且对于一切 $x\geqslant 0, y\geqslant 0, g(x+y)\leqslant g(x)+g(y)$.) 然而, 对于在 $(\Omega,\mathscr{F},\mathbf{P})$ 上的一切随机变量的空间*不存在*这样的距离 $d(\xi,\eta)$, 使得当且仅当 ξ_n 依概率 1 收敛于 ξ 时 $d(\xi_n,\xi)\to 0$. (这很容易证明: 只需考虑依概率收敛, 但是不依概率 1 收敛的随机变量序列 $\xi_n, n\geqslant 1$.) 换句话说, *依概率 1 收敛不可度量化*. (见第二章 §10 中练习题 11 和 2 的命题.)

这一节的目的, 是在指出度量 ($L(P,\widetilde{P})$ 和 $\|P-\widetilde{P}\|^*_{BL}$) 的同时, 在测度 $\mathscr{P}(E)$ 的空间中, 建立*弱收敛的可度量性*:

$$P_n \xrightarrow{w} P \Leftrightarrow L(P_n,P)\to 0 \Leftrightarrow \|P_n-P\|^*_{BL}\to 0. \tag{1}$$

2. 列维 – 普罗霍罗夫度量 $L(P,\widetilde{P})$ 设

$$\rho(x,A)=\inf\{\rho(x,y):y\in A\},$$
$$A^{\varepsilon}=\{x\in E:\rho(x,A)<\varepsilon\}, A\in\mathscr{E}.$$

对于任意两测度 $P,\widetilde{P}\in\mathscr{P}(\mathscr{E})$, 设

$$\sigma(P,\widetilde{P})=\inf\{\varepsilon>0:P(F)\leqslant\widetilde{P}(F^{\varepsilon})+\varepsilon, \text{对于一切闭集} F\in\mathscr{E}\}, \tag{2}$$

和

$$L(P,\widetilde{P})=\max\{\sigma(P,\widetilde{P}),\sigma(\widetilde{P},P)\}. \tag{3}$$

下面的引理证明, 这样定义的所谓*列维 – 普罗霍罗夫度量*函数 $L(P,\widetilde{P}), P,\widetilde{P}\in\mathscr{P}(\mathscr{E})$, 确实是度量.

引理 1 函数 $L(P, \widetilde{P})$ 具有距离的性质:

a) $L(P, \widetilde{P}) = L(\widetilde{P}, P)(= \sigma(P, \widetilde{P}) = \sigma(\widetilde{P}, P))$,

b) $L(P, \widetilde{P}) \leqslant L(P, \widehat{P}) + \sigma(\widehat{P}, \widetilde{P})$,

c) $L(P, \widetilde{P}) = 0$, 当且仅当 $P = \widetilde{P}$.

证明 a) 只需证明, 对于 $\alpha > 0, \beta > 0$,

$$\text{“}P(F) \leqslant \widetilde{P}(F^\alpha) + \beta, \text{ 对于一切闭集 } F \in \mathscr{E}\text{”} \tag{4}$$

当且仅当

$$\text{“}\widetilde{P}(F) \leqslant P(F^\alpha) + \beta, \text{ 对于一切闭集 } F \in \mathscr{E}\text{”}. \tag{5}$$

设 T 是 $\mathscr{E}$ 中的闭集. 那么,T^α 是开集, 则不难验证 $T \subseteq E\backslash(E\backslash T^\alpha)^\alpha$. 如果 (4) 式成立, 那么, 特别

$$P(E\backslash T^\alpha) \leqslant \widetilde{P}((E\backslash T^\alpha)^\alpha) + \beta,$$

从而

$$\widetilde{P}(T) \leqslant \widetilde{P}(E\backslash(E\backslash T^\alpha)^\alpha) \leqslant P(T^\alpha) + \beta,$$

因此, 说明 (4) 和 (5) 两式等价. 由此可见

$$\sigma(P, \widetilde{P}) = \sigma(\widetilde{P}, P), \tag{6}$$

于是,

$$L(P, \widetilde{P}) = \sigma(P, \widetilde{P}) = \sigma(\widetilde{P}, P) = L(\widetilde{P}, P). \tag{7}$$

b) 设 $L(P, \hat{P}) < \delta_1, L(\hat{P}, \widetilde{P}) < \delta_2$. 那么, 对于每一个闭集 $F \in \mathscr{E}$, 有

$$\widetilde{P}(F) \leqslant \widehat{P}(F^{\delta_2}) + \delta_2 \leqslant P((F^{\delta_2})^{\delta_1}) + \delta_1 + \delta_2 \leqslant P(F^{\delta_1+\delta_2}) + \delta_1 + \delta_2,$$

于是,$L(P, \widetilde{P}) \leqslant \delta_1 + \delta_2$. 由此可见

$$L(P, \widetilde{P}) \leqslant L(P, \widehat{P}) + L(\widehat{P}, \widetilde{P}).$$

c) 如果 $L(P, \widetilde{P}) = 0$, 那么, 对于每一个闭集 $F \in \mathscr{E}$ 和任意 $\alpha > 0$, 有

$$P(F) \leqslant \widetilde{P}(F^\alpha) + \alpha. \tag{8}$$

由于 $F^\alpha \downarrow F, \alpha \downarrow 0$, 则由 (8) 式当 $\alpha \downarrow 0$ 时求极限, 得 $P(F) \leqslant \widetilde{P}(F)$; 由对称性, 有 $\widetilde{P}(F) \leqslant P(F)$. 这样, 对于一切闭集 $F \in \mathscr{E}$, 有 $P(F) = \widetilde{P}(F)$. 对于每一个博雷尔集 $A \in \mathscr{E}$ 和任意 $\varepsilon > 0$, 存在这样的开集 $G_\varepsilon \supseteq A$ 和闭集 $F_\varepsilon \subseteq A$, 且 $P(G_\varepsilon \setminus F_\varepsilon) \leqslant \varepsilon$. 由此可见, 度量空间 $(E, \mathscr{E}, \rho)$ 上的任何概率测度 P, 完全决定于 P 在闭集上的值. 从而, 由对于一切闭集 $F \in \mathscr{E}$, 有 $\widetilde{P}(F) = P(F)$. 于是, 对于一切博雷尔集 $A \in \mathscr{E}$, 有 $\widetilde{P}(A) = P(A)$. □

定理 1 列维－普罗霍罗夫度量 $L(P,\widetilde{P})$, 可以使弱收敛度量化:

$$L(P_n,P)\to 0\Leftrightarrow P_n\xrightarrow{w}P. \tag{9}$$

证明 $(\Rightarrow)$. 设 $L(P_n,P)\to 0, n\to\infty$. 那么, 对于任意固定的闭集 $F\in\mathscr{E}$ 和任何 $\varepsilon>0$, 根据 (2) 式和引理 1 的命题 a), 有

$$\overline{\lim_n} P_n(F)\leqslant P(F^{\varepsilon})+\varepsilon. \tag{10}$$

在上式中令 $\varepsilon\downarrow 0$, 得

$$\overline{\lim_n} P_n(F)\leqslant P(F).$$

根据 §1 的定理 1, 可见

$$P_n\xrightarrow{w}P. \tag{11}$$

$(\Leftarrow)$. 蕴涵关系 $(\Leftarrow)$ 的证明, 基于一系列深刻而有益的事实, 它既进一步揭示弱收敛本身的内容, 又阐明其构建方法和研究收敛“速度”的方法.

这样, 设 $P_n\xrightarrow{w}P$. 这表明对于任意连续的有界函数 $f=f(x)$, 有

$$\int_E f(x)P_n(\mathrm{d}x)\to\int_E f(x)P(\mathrm{d}x). \tag{12}$$

现在假设 $\mathscr{G}$ 是某一等阶梯函数 $g=g(x)$ 类 (如果对于一切 $g\in\mathscr{G}$, 有 $\rho(x,y)<\delta$, 则对于任意 $\varepsilon>0$ 存在 $\delta>0$, 使 $|g(y)-g(x)|<\varepsilon$), 使得对于同一常数 $C>0$,(对于一切 $x\in E$ 和 $g\in\mathscr{G}$)$|g(x)|\leqslant C$. 根据 §8 定理 3, 对于函数类 $\mathscr{G}$, 性质 (12) 的如下条件成立:

$$P_n\xrightarrow{w}P\Rightarrow\sup_{g\in\mathscr{G}}\left|\int_E g(x)P_n(\mathrm{d}x)-\int_E g(x)P(\mathrm{d}x)\right|\to 0. \tag{13}$$

对于任意 $A\in\mathscr{E}$ 和 $\varepsilon>0$, (如同 §1 定理 1) 记

$$f_A^{\varepsilon}(x)=\left[1-\frac{\rho(x,A)}{\varepsilon}\right]^{+}. \tag{14}$$

显然

$$I_A(x)\leqslant f_A^{\varepsilon}(x)\leqslant I_{A^{\varepsilon}}(x), \tag{15}$$

且

$$|f_A^{\varepsilon}(x)-f_A^{\varepsilon}(y)|\leqslant\varepsilon^{-1}|\rho(x,A)-\rho(y,A)|\leqslant\varepsilon^{-1}\rho(x,y).$$

这样, 对于类 $\mathscr{G}^{\varepsilon}=\{f_A^{\varepsilon}(x),A\in\mathscr{E}\}$, 有 (13) 式. 因此,

$$\Delta_n\equiv\sup_{A\in\mathscr{E}}\left|\int_E f_A^{\varepsilon}(x)P_n(\mathrm{d}x)-\int_E f_A^{\varepsilon}(x)P(\mathrm{d}x)\right|\to 0, n\to\infty. \tag{16}$$

由此及 (15) 式, 可见对于任意闭集 $A \in \mathscr{E}$ 和 $\varepsilon > 0$, 有

$$P(A^{\varepsilon}) \geqslant \int_E f_A^{\varepsilon}(x)\mathrm{d}P \geqslant \int_E f_A^{\varepsilon}(x)\mathrm{d}P_n - \Delta_n \geqslant P_n(A) - \Delta_n. \tag{17}$$

选择 $n(\varepsilon)$ 使对于一切 $n \geqslant n(\varepsilon)$, 有 $\Delta_n \leqslant \varepsilon$. 那么, 由 (17) 式, 对于 $n \geqslant n(\varepsilon)$, 有

$$P(A^{\varepsilon}) \geqslant P_n(A) - \varepsilon. \tag{18}$$

由此及定义 (2),(3) 式可见, 当 $n \geqslant n(\varepsilon)$ 时, 有 $L(P_n, P) \leqslant \varepsilon$. 于是,

$$P_n \xrightarrow{w} P \Rightarrow \Delta_n \to 0 \Rightarrow L(P_n, P) \to 0.$$

定理 (精确到 (13) 式的命题) 得证. □

3. 度量 $\|P - \widetilde{P}\|_{BL}^*$ 以 BL 表示一切连续有界函数 $f = f(x), x \in E$ 的集合, 其中 $\|f\|_\infty = \sup_x |f(x)| < \infty$, 而且每一个函数 $f = f(x)$ 都满足利普希茨条件:

$$\|f\|_L = \sup_{x \neq y} \frac{|f(x) - f(y)|}{\rho(x, y)} < \infty.$$

设 $\|f\|_{BL} = \|f\|_\infty + \|f\|_L$. 以 $\|\cdot\|_{BL}$ 为范数的空间 BL 是巴拿赫空间.

定义度量 $\|P - \widetilde{P}\|_{BL}^*$, 其中

$$\|P - \widetilde{P}\|_{BL}^* = \sup_{f \in BL} \left\{ \left| \int f\mathrm{d}(P - \widetilde{P}) \right| : \|f\|_{BL} \leqslant 1 \right\}. \tag{19}$$

(可以验证, $\|P - \widetilde{P}\|_{BL}^*$ 事实上满足对于度量提出的一切要求; 练习题 2.)

定理 2 *度量 $\|P - \widetilde{P}\|_{BL}^*$ 可以使弱收敛度量化:*

$$\|P_n - P\|_{BL}^* \to 0 \Leftrightarrow P_n \xrightarrow{w} P.$$

证明 由 (13) 式立即得蕴涵关系 ($\Leftarrow$). 为证明 ($\Rightarrow$) 只需验证, 在弱收敛的定义中, $P_n \xrightarrow{w} P$ 就是对于任意有界连续函数 $f = f(x)$, 满足性质 (12) 式, 故只需局限于考虑满足利普希茨条件的有界函数类. 换句话说, 如果能证明下面的结果 (引理), 则蕴涵关系 ($\Rightarrow$) 将得到证明.

引理 2 *弱收敛 $P_n \xrightarrow{w} P$ 成立, 当且仅当性质 (12) 对于 BL 类的任何函数 $f = f(x)$ 成立.*

证明 引理的一个方面的证明显然. 现在考虑由 (14) 式定义的函数 $f_A^{\varepsilon} = f_A^{\varepsilon}(x)$. 在证明定理 1 时已经证明, 对于任意 $\varepsilon > 0, \mathscr{G}^{\varepsilon} = \{f_A^{\varepsilon}(x), A \in \mathscr{E}\} \subseteq BL$. 如果现在分析 §1 的定理 1 中蕴涵关系 (I)⇒(Ⅱ) 的证明, 就会发现, 在证明中实际上并非对于一切有界连续函数用到性质 (12), 而只对于类 $\mathscr{G}^{\varepsilon}(\varepsilon > 0)$ 中的函数用到. 由于 $\mathscr{G}^{\varepsilon} \subseteq BL, \varepsilon > 0$, 则显然由性质 (12) 对于类 BL 中的函数成立, 可见 §1 定理 1 中命题 Ⅱ 的结论成立, 而后者 (仍然因为 §1 定理 1) 等价于弱收敛 $P_n \xrightarrow{w} P$. □

注. 定理 2 的结论可以由定理 1 推出 (反之也一样), 假如利用 (对于可分度量空间 $(E,\mathscr{E},\rho)$ 成立的) $L(P,\widetilde{P})$ 和 $\|P-\widetilde{P}\|_{BL}^*$ 之间的如下不等式:

$$\|P-\widetilde{P}\|_{BL}^* \leqslant 2L(P,\widetilde{P}), \tag{20}$$

$$\varphi(L(P,\widetilde{P})) \leqslant \|P-\widetilde{P}\|_{BL}^*, \quad 其中 \quad \varphi(x)=\frac{2x^2}{2+x}. \tag{21}$$

注意到, 对于 $x \geqslant 0, 0 \leqslant \varphi(x) \leqslant 2/3$, 当且仅当 $x \leqslant 1$, 且对于 $0 \leqslant x \leqslant 1, \varphi(x) \geqslant 2x^2/3$; 由 (20) 和 (21) 式, 可见: 如果 $L(P,\widetilde{P}) \leqslant 1$ 或 $\|P-\widetilde{P}\|_{BL}^* \leqslant 2/3$, 则

$$\frac{2}{3}L^2(P,\widetilde{P}) \leqslant \|P-\widetilde{P}\|_{BL}^* \leqslant 2L(P,\widetilde{P}). \tag{22}$$

4. 练习题

1. 证明, 当 $E=\mathbb{R}$ 时概率分布 P 和 $\widetilde{P}$ 之间的列维 - 普罗霍罗夫度量 $L(P,\widetilde{P})$, 不小于对应于 P 和 $\widetilde{P}$ 的分布函数 F 和 $\widetilde{F}$ 之间的列维距离 $L(F,\widetilde{F})$ (见 §1 练习题 4). 举例说明这些度量间的严格不等式成立.

2. 证明公式 (19) 在决定空间BL中的度量.

3. 证明不等式 (20)(21) 和 (22).

4. 设 $F=F(x)$ 和 $G=G(x)$ 是两个分布函数, P_c 和 Q_c 是它们与直线 $x+y=c$ 的交点. 证明列维距离 $L(F,\widetilde{F})$ (见 §1 练习题 4)

$$L(F,G)=\sup_c \frac{\overline{P_cQ_c}}{\sqrt{2}},$$

其中 $\overline{P_cQ_c}$ 表示点 P_c 和 Q_c 之间的线段的长度.

5. 证明, 一切分布函数的集合, 关于列维距离的是完备空间.

§8. 关于测度的弱收敛与随机元的几乎处处收敛的联系 ("一个概率空间的方法")

1. 随机元收敛性的定义 假设在概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 上给定随机元 $X=X(\omega)$, $X_n=X_n(x), n \geqslant 1$, 而 $(E,\mathscr{E},\rho)$ 是随机元的值空间; 参见第二章 §5. 记 P 和 P_n 是 X 和 X_n 的概率分布, 即设

$$P(A)=\mathbf{P}\{\omega: X(\omega)\in A\}, \quad P_n(A)=\mathbf{P}\{\omega: X_n(\omega)\in A\}, A\in\mathscr{E}.$$

推广随机变量依分布收敛的概念 (见第二章 §10), 引进下面的定义.

定义 1 称随机元序列 $X_n, n \geqslant 1$ 为*依分布收敛或依分布律收敛的* (记号: $X_n \xrightarrow{\mathscr{D}} X$, 或 $X_n \xrightarrow{\mathscr{L}} X$, 或 $X_n \xrightarrow{Law} X$), 如果 $P_n \xrightarrow{w} P$.

仿照随机变量的依概率收敛和依概率 1 收敛 (第二章 §10), 下列定义是自然的.

定义 2 称随机元序列$X_n, n \geqslant 1$依概率收敛于 X ($X_n \xrightarrow{\mathbf{P}} X$), 如果

$$\mathbf{P}\{\omega : \rho(X_n(\omega), X(\omega)) \geqslant \varepsilon\} \to 0, n \to \infty. \tag{1}$$

定义 3 称随机元序列 $X_n, n \geqslant 1$, 依概率 1 (几乎必然或几乎处处) 收敛于 $X(X_n \xrightarrow{\text{a.c.}} X, X_n \xrightarrow{\text{a.e.}} X)$, 如果 $\rho(X_n(\omega), X(\omega)) \xrightarrow{\text{a.c.}} 0, n \to \infty$.

注 1 当然, 只有在 $\rho(X_n(\omega), X(\omega))$ 作为 $\omega \in \Omega$ 的函数是随机变量时, 即当 $\rho(X_n(\omega), X(\omega))$ 为 $\mathscr{F}$- 可测时, 最后两个定义才有意义. 当空间 $(E, \mathscr{E}, \rho)$ 可分时, 这当然成立 (练习题 1).

注 2 关于定义 2 需要指出, 引进的依概率收敛性, 可以由下面的 (定义在 $(\Omega, \mathscr{F}, \mathbf{P})$ 上、取值于 E 的; 练习题 2) 随机元 X 和 Y 之间的*范基* (Fan Ky) *度量* (练习题 2)

$$d_{\mathbf{P}}(X, Y) = \inf\{\varepsilon > 0 : \mathbf{P}\{\rho(X(\omega), Y(\omega)) \geqslant \varepsilon\} \leqslant \varepsilon\} \tag{2}$$

来度量化.

注 3 如果依概率收敛和依概率 1 收敛的定义, 要求随机元定义在同一概率空间上, 则按分布收敛的定义 $X_n \xrightarrow{\mathscr{D}} X$, 只与分布的收敛性有关, 从而可以认为 $X(\omega), X_1(\omega), X_2(\omega), \cdots$ 在同一空间 E 上取值, 不过可能给定在"*各自的*"概率空间 $(\Omega, \mathscr{F}, \mathbf{P}), (\Omega_1, \mathscr{F}_1, \mathbf{P}_1), (\Omega_2, \mathscr{F}_2, \mathbf{P}_2), \cdots$ 上. 然而不失普遍性, 如果作为上述各空间的直积, 并且定义 $X(\omega, \omega_1, \omega_2, \cdots) = X(\omega), X_1(\omega, \omega_1, \omega_2, \cdots) = X_1(\omega_1), \cdots$, 则总可以认为它们定义在同一个概率空间上.

2. 按分布相等的随机元 根据定义 1 和在勒贝格积分中变量替换的定理 (第二章 §6 定理 7), 对于任意有界连续函数 $f = f(x), x \in E$, 有

$$X_n \xrightarrow{\mathscr{D}} X \Leftrightarrow \mathbf{E}f(X_n) \to f(X). \tag{3}$$

由 (3) 式可见, 根据勒贝格控制收敛定理 (第二章 §6 定理 3), 由收敛性 $X_n \xrightarrow{\text{a.c.}} X$, 立即可以得到 $X_n \xrightarrow{\mathscr{D}} X$. 因此, 显然对于 X 和 X_n 是随机变量的情形, 有同样的结果 (第二章 §10 定理 2). 更意想不到的是, 在某种意义上有相反的结果. 下面就给出其确切的提法, 然后讨论其应用.

定义 4 设随机元 $X = X(\omega')$ 和 $Y = Y(\omega'')$, 分别定义在概率空间 $(\Omega', \mathscr{F}', \mathbf{P}')$ 和 $(\Omega'', \mathscr{F}'', \mathbf{P}'')$ 上, 取值于同一空间 E. 如果 $X = X(\omega')$ 和 $Y = Y(\omega'')$ 有相同的概率分布, 则称之为*按分布为相等的* (*等价的或相合的*; 记作 $X \overset{\mathscr{D}}{=} Y$).

定理 1 *假设 $(E, \mathscr{E}, \rho)$ 是可分度量空间.*

1. 设随机元 $X, X_n (n \geqslant 1)$, 定义在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上, 取值于 E 中, 且 $X_n \xrightarrow{\mathscr{D}} X$. 那么, 存在概率空间 $(\Omega^*, \mathscr{F}^*, \mathbf{P}^*)$ 和定义在该空间上且取值于 E 的随机元 $X^*, X_n^* (n \geqslant 1)$, 使

$$X_n^* \xrightarrow{\text{a.c.}} X^*,$$

且

$$X^* \overset{\mathscr{D}}{=} X, \quad X_n^* \overset{\mathscr{D}}{=} X_n, \quad n \geqslant 1.$$

2. 设 $P, P_n(n \geqslant 1)$ 是 $(E, \mathscr{E}, \rho)$ 上的概率测度, 且 $P_n \xrightarrow{w} P$. 那么, 存在概率空间 $(\Omega^*, \mathscr{F}^*, \mathbf{P}^*)$ 和定义在该空间上且取值于 E 的随机元 $X^*, X_n^*(n \geqslant 1)$, 使

$$X_n^* \xrightarrow{\text{a.c.}} X^*,$$

且

$$P^* = P, \quad P_n^* = P_n, \quad n \geqslant 1,$$

其中 P^* 和 P_n^* 是 X^* 和 X_n^* 的概率分布.

在证明定理之前, 我们指出: 第一, 只需证明命题 2; 因为, 如果把 P 和 P_n 取作 X 和 X_n 的概率分布, 则命题 1 可以由命题 2 得到. 第二, 需要指出, 这一定理最一般情形的证明技术上相当复杂. 正因为如此, 我们仅对 $E = \mathbb{R}$ 的情形证明该定理. 这一证明相当简明易懂, 并且为所求对象给出了简单描述性构造. (美中不足的是, 这一构造在一般情形下, 甚至对于 $E = \mathbb{R}^2$ 的情形, 并不 "适用".)

定理 1 的证明 ($E = \mathbb{R}$ 的情形). 设 $F = F(x)$ 和 $F_n = F_n(x)$ 是对应于 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的测度 P 和 P_n 的分布函数; 而是 $Q = Q(u)$ 与函数 $F = F(x)$ 相联系的分位函数, 唯一地决定于如下公式

$$Q(u) = \inf\{x : F(x) \geqslant u\}, \quad 0 < u < 1. \tag{4}$$

不难验证,

$$F(x) \geqslant u \Leftrightarrow Q(u) \leqslant x. \tag{5}$$

现在考虑 $\Omega^* = (0,1), \mathscr{F}^*(x) = \mathscr{B}(0,1)$, 而 $\mathbf{P}^*$ 是勒贝格测度: $\mathbf{P}^*(d\omega^*) = d\omega^*$. 设 $X^*(\omega^*) = Q(\omega^*), \omega^* \in \Omega^*$. 那么,

$$\mathbf{P}^*\{\omega^* : X^*(\omega^*) \leqslant x\} = \mathbf{P}^*\{\omega^* : Q(\omega^*) \leqslant x\} = \mathbf{P}^*\{\omega^* : \omega^* \leqslant F(x)\} = F(x),$$

即所构造的随机变量 $X^*(\omega^*) = Q(\omega^*)$ 的分布恰好是 P. 类似地, 随机变量 $X_n^*(\omega^*) = Q_n(\omega^*)$ 的分布恰好是 P_n.

其次, 并不复杂地可以证明, 由在极限函数 $F = F(x)$ 的每个连续点上, $F_n(x)$ 收敛于 $F(x)$ (对于 $E = \mathbb{R}$ 的情形, 等价于 $P_n \xrightarrow{w} P$; 见 §1 的定理 1), 可见分位函数序列 $Q_n(u), n \geqslant 1$, 在极限函数 $Q = Q(u)$ 的每个连续点上, 也收敛于 $Q(u)$. 由于函数 $Q = Q(u), u \in (0,1)$, 最多有可数个间断点, 则其勒贝格测度 $\mathbf{P}^*$ 等于 0. 因而

$$X_n^*(\omega^*) = Q_n(\omega^*) \xrightarrow{\text{a.c.}} X^*(\omega^*) = Q(\omega^*).$$

定理 1 (对于 $E = \mathbb{R}$ 的情形) 得证. □

定理 1 所描绘的, 由给定的随机元 X 和 X_n, 向定义在同一概率空间上的新随机元 X^* 和 X_n^* 的过渡, 说明了本节的标题中的名称 "一个概率空间的方法" 的含义.

现在利用这一方法, 可以比较容易地证明一系列的命题.

3. 一个概率空间方法的应用 假设 $X, X_n(n \geqslant 1)$ 是定义在同一概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上, 在可分度量空间 $(E, \mathscr{E}, \rho)$ 中取值的随机元, 且 $X_n \xrightarrow{\mathscr{D}} X$. 设 $h = h(x), x \in E$, 是 $(E, \mathscr{E}, \rho)$ 到另外一个可分度量空间 $(E', \mathscr{E}', \rho')$ 上的可测影射. 在概率论和数理统计中, 常遇到这样的问题: 在$h = h(x)$ 满足什么条件时, 由收敛性 $X_n \xrightarrow{\mathscr{D}} X$ 可以得出收敛性 $h(X_n) \xrightarrow{\mathscr{D}} h(X)$.

例如, 假设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量, 且 $\mathbf{E}\xi_1 = m, \mathbf{D}\xi_1 = \sigma^2 > 0$, 而 $\overline{X}_n = (\xi_1 + \cdots + \xi_n)/n$. 由中心极限定理, 可见

$$\frac{\sqrt{n}(\overline{X}_n - m)}{\sigma} \xrightarrow{d} N(0,1).$$

问对于什么函数 $h = h(x)$ 可以保障

$$h\left(\frac{\sqrt{n}(\overline{X}_n - m)}{\sigma}\right) \xrightarrow{d} h(N(0,1))?$$

(对于连续函数 $h = h(x)$ 肯定可运用著名的曼 - 沃尔德 [H. B. Mann – A. Wald] 定理, 因此立即可得:

$$\frac{n(\overline{X}_n - m)^2}{\sigma^2} \xrightarrow{d} \chi_1^2,$$

其中 χ_1^2 是自由度为 1 的 χ^2 分布随机变量; 见第一章 §3 表 1 – 2.)

另一例子. 如果 $X = X(t,\omega), X_n = X_n(t,\omega), t \in T$, 是随机过程 (见第二章 §5), 而

$$h(X) = \sup_{t \in T} |X(t,\omega)|, h(X_n) = \sup_{t \in T} |X_n(t,\omega)|,$$

则上面所提问题表示: 在什么条件下由按分布过程收敛 $X_n \xrightarrow{\mathscr{D}} X$, 可以得出其上确界按分布收敛 $h(X_n) \xrightarrow{\mathscr{D}} h(X)$?

保障蕴涵关系

$$X_n \xrightarrow{\mathscr{D}} X \Rightarrow h(X_n) \xrightarrow{\mathscr{D}} h(X)$$

的条件之一是, 影射 $h = h(x)$ 连续. 事实上, 如果 $f = f(x')$ 在 E' 上是有界连续函数, 则 $f = f(h(x))$ 在 E 上也是有界连续函数. 从而

$$X_n \xrightarrow{\mathscr{D}} X \Rightarrow \mathbf{E}f(h(X_n)) \to \mathbf{E}f(h(X)).$$

下面的定理表明, 考虑到极限随机元 X 的性质, 对于函数 $h = h(x)$ 连续性的要求可以进行一定减弱.

记 $\Delta_h = \{x \in E : h(x)$ 在点 x 处不 ρ- 连续 $\}$, 换句话说, 设 Δ_h 是函数 $h = h(x)$ 的间断点的集合. 注意 $\Delta_h \in \mathscr{E}$ (练习题 4).

定理 2 1. 设 $(E,\mathscr{E},\rho)$ 和 $(E',\mathscr{E}',\rho')$ 是可分度量空间, 且 $X_n \xrightarrow{\mathscr{D}} X$, 而影射 $h=h(x), x\in E$, 满足条件:

$$\mathbf{P}\{\omega : x \in \Delta_h\} = 0. \tag{6}$$

那么,$h(X_n) \xrightarrow{\mathscr{D}} h(X)$.

2. 设 $P, P_n(n\geqslant 1)$ 是可分度量空间 $(E,\mathscr{E},\rho)$ 上的概率测度, 且 $P_n \xrightarrow{w} P$, 而 $h=h(x), x\in E$, 是 $(E,\mathscr{E},\rho)$ 到可分度量空间 $(E',\mathscr{E}',\rho')$ 的可测影射. 假设

$$P\{x : x\in\Delta_h\}=0.$$

那么,$P_n^h \xrightarrow{w} P^h$, 其中 $P_n^h(A)=P_n\{h(x)\in A\}, P^h(A)=P\{h(x)\in A\}, A\in\mathscr{E}'$.

证明 像定理 1 一样, 例如, 只需证明命题 1.

设 X^* 和 $X_n^*(n\geqslant 1)$ 是用 "一个概率空间的方法" 建立的随机元, 且满足 $X^* \stackrel{\mathscr{D}}{=} X, X_n^* \stackrel{\mathscr{D}}{=} X_n(n\geqslant 1), X_n^* \xrightarrow{\text{a.c.}} X$. 记

$$A^*=\{\omega^* : \rho(X_n^*, X^*) \nrightarrow 0\}, \quad B^*=\{\omega^* : X^*(\omega^*)\in\Delta_n\}.$$

那么,$\mathbf{P}^*(A^*\cup B^*)=0$ 且对于 $\omega^*\notin A^*\cup B^*$,

$$h(X_n^*(\omega^*)) \to h(X^*(\omega^*)),$$

而这说明 $h(X_n^*) \xrightarrow{\text{a.c.}} h(X^*)$. 在第 1 小节已经指出, 由此可见 $h(X_n^*) \xrightarrow{\mathscr{D}} h(X^*)$. 由于 $h(X_n^*) \stackrel{\mathscr{D}}{=} h(X_n), h(X^*) \stackrel{\mathscr{D}}{=} h(X)$; 可见 $h(X_n) \xrightarrow{\mathscr{D}} h(X)$. □

4. 完成 §7 中 (13) 式的证明 在 §7 证明定理 1 的蕴涵关系 (⇐) 时, 曾经用到性质 (13). 现在进行 (13) 式的证明, 仍然借助 "一个概率空间的方法".

设 $(E,\mathscr{E},\rho)$ 是可分度量空间, $\mathscr{G}$ 是等阶梯连续函数 $g=g(x)$ 类, 满足性质: 对于一切 $x\in E, g\in\mathscr{G}$ 和某个常数 C, 有 $|g(x)|\leqslant C$.

定理 3 设 $P, P_n(n\geqslant 1)$ 是 $(E,\mathscr{E},\rho)$ 上的概率测度, 且 $P_n \xrightarrow{w} P$, 那么

$$\sup_{g\in\mathscr{G}}\left|\int_E g(x)P_n(\mathrm{d}x)-\int_E g(x)P(\mathrm{d}x)\right| \to 0, \quad n\to\infty. \tag{7}$$

证明 设 (7) 式成立, 则存在 $a>0$ 和函数 $g_1, g_2, \cdots \in\mathscr{G}$, 使对于无限多个$n$, 有

$$\left|\int_E g_n(x)P_n(\mathrm{d}x)-\int_E g_n(x)P(\mathrm{d}x)\right| \geqslant a>0. \tag{8}$$

将 "一个概率空间的方法" 用于随机元 X^* 和 X_n^* (见定理 1), 可以将 (8) 式写成: 对于无限多个 n,

$$|\mathbf{E}^* g_n(X_n^*)-\mathbf{E}^* g_n(X^*)| \geqslant a>0. \tag{9}$$

由于类 $\mathscr{G}$ 的性质, 对于任意 $\varepsilon>0$, 存在 $\delta>0$, 当 $\rho(x,y)<\delta$ 时, 使对于一切 $g\in\mathscr{G}$, 有 $|g(y)-g(x)|<\varepsilon$. 此外, 对于一切 $x\in E, g\in\mathscr{G}$, 有 $|g(x)|\leqslant C$. 所以

$$\begin{aligned}&|\mathbf{E}^*g_n(X_n^*)-\mathbf{E}^*g_n(X^*)|\\ \leqslant{}& \mathbf{E}^*\{|g_n(X_n^*)-g_n(X^*)|;\rho(X_n^*,X^*)>\delta\}\\ &+\mathbf{E}^*\{|g_n(X_n^*)-g_n(X^*)|;\rho(X_n^*,X^*)\leqslant\delta\}\\ \leqslant{}& 2C\mathbf{P}^*\{\rho(X_n^*,X^*)>\delta\}+\varepsilon.\end{aligned}$$

由于 $X_n^*\xrightarrow{\text{a.c.}}X$, 故 $\mathbf{P}^*\{\rho(X_n^*,X^*)>\delta\}\to 0, n\to\infty$. 于是, 由于 $\varepsilon>0$ 任意性, 可见

$$\overline{\lim_n}|\mathbf{E}^*g_n(X_n^*)-\mathbf{E}^*g_n(X^*)|=0,$$

而这与 (9) 式矛盾. □

5. 列维 – 普罗霍罗夫度量值的上估计　在这一小节把用于定理 1 的 "一个概率空间的方法" 的思想, 用来从上侧估计可分度量空间 $(E,\mathscr{E},\rho)$ 上两个概率分布之间的列维 – 普罗霍罗夫度量的值 $L(P,\widetilde{P})$.

定理 4　*对于任意两个测度 P 和 $\widetilde{P}$, 存在概率空间 $(\Omega^*,\mathscr{F}^*,\mathbf{P}^*)$, 以及定义在该空间上、取值于 E 的随机元 X 和 $\widetilde{X}$, 而且 X 和 $\widetilde{X}$ 的分布恰好分别为 P 和 $\widetilde{P}$, 并且满足:*

$$L(P,\widetilde{P})\leqslant d_{\mathbf{P}^*}(X,\widetilde{X})=\inf\{\varepsilon>0;\mathbf{P}^*\{\rho(X,\widetilde{X})\geqslant\varepsilon\}\leqslant\varepsilon\}. \tag{10}$$

证明　根据定理 1, 确实存在概率空间 $(\Omega^*,\mathscr{F}^*,\mathbf{P}^*)$ 定义在该空间上、取值于 E 的随机元 X 和 $\widetilde{X}$, 使 $\mathbf{P}^*\{X\in A\}=P(A), \mathbf{P}^*\{\widetilde{X}\in A\}=\widetilde{P}(A), A\in\mathscr{E}$.

设 $\varepsilon>0$, 使

$$\mathbf{P}^*\{\rho(X,\widetilde{X})\geqslant\varepsilon\}\leqslant\varepsilon. \tag{11}$$

那么, 对于任意 $A\in\mathscr{E}$,

$$\begin{aligned}\widetilde{P}(A)=\mathbf{P}^*\{\widetilde{X}\in A\}&=\mathbf{P}^*\{\widetilde{X}\in A, X\in A^\varepsilon\}+\mathbf{P}^*\{\widetilde{X}\in A, X\notin A^\varepsilon\}\\ &\leqslant\mathbf{P}^*\{X\in A^\varepsilon\}+\mathbf{P}^*\{\rho(X,\widetilde{X})\geqslant\varepsilon\}\leqslant P(A^\varepsilon)+\varepsilon,\end{aligned}$$

其中 $A^\varepsilon=\{x\in E:\rho(x,A)<\varepsilon\}$. 因此, 根据列维 – 普罗霍罗夫度量的定义 (§7 第 2 小节),

$$L(P,\widetilde{P})\leqslant\varepsilon. \tag{12}$$

由 (11) 和 (12) 式, 并对 $\varepsilon>0$ 求 inf, 得所需要求的 (10) 式. □

系　*设 X 和 $\widetilde{X}$ 是概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 上、取值于 E 的随机元, 其概率分布为 P_X 和 $P_{\widetilde{X}}$, 那么,*

$$L(P_X,P_{\widetilde{X}})\leqslant d_{\mathbf{P}}(X,\widetilde{X}).$$

注 1 所作的证明显示, 事实上 (10) 式总是正确的, 只要在某一概率空间 $(\Omega^*, \mathscr{F}^*, \mathbf{P}^*)$ 上可以指出取值于 E 的随机元 X 和 $\widetilde{X}$, 使其概率分布相应为 P 和 $\widetilde{P}$, 而且对于 X 和 $\widetilde{X}$, 有 $\{\omega^*: \rho(X(\omega^*), \widetilde{X}(\omega^*)) \geqslant \varepsilon\} \in \mathscr{F}^*, \varepsilon > 0$. 因而, 估计式 (10) 的质量本质上依赖于, 根据测度 P 和 $\widetilde{P}$ 建立的对象 $(\Omega^*, \mathscr{F}^*, \mathbf{P}^*)$ 和 $X, \widetilde{X}$ 如何. (由 P 和 $\widetilde{P}$ 建立 $\Omega^*, \mathscr{F}^*, \mathbf{P}^*$ 和 $X, \widetilde{X}$ 的过程, 称做**联结**或**匹配** —— 源于英语单词 coupling.) 例如, 可以取 "联结 - 测度" $\mathbf{P}^*$ 等于测度 P 和 $\widetilde{P}$ 的直积: $\mathbf{P}^* = P \times \widetilde{P}$, 然而这样的选择通常不会得到 (10) 式的好估计.

注 2 自然提出问题, (10) 式何时为等式. 为此, 我们 (不加证明) 给出如下结果: *设 P 和 $\widetilde{P}$ 是可分度量空间 $(E, \mathscr{E}, \rho)$ 上的两个概率测度, 存在这样的 $(\Omega^*, \mathscr{F}^*, \mathbf{P}^*)$ 和 $X, \widetilde{X}$, 使*

$$L(P, \widetilde{P}) = d_{\mathbf{P}^*}(X, \widetilde{X}) = \inf\{\varepsilon > 0 : \mathbf{P}^*\{\rho(X, \widetilde{X}) \geqslant \varepsilon\} \leqslant \varepsilon\}.$$

6. 练习题

1. 假设概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 是某一可分度量空间, 而 $X(\omega)$ 和 $Y(\omega)$ 是定义在该空间上的任意随机元, 证明实函数 $\rho(X(\omega), Y(\omega))$ 是随机变量.

2. 证明由 (2) 式定义的函数 $d_{\mathbf{P}}(X, Y)$, 是取值于 E 的随机元空间中的度量.

3. 证明 (5) 式的正确性.

4. 证明集合 $\Delta_h = \{x \in E : h(x)\text{在点}x\text{关于距离}\rho\text{不连续}\} \in \mathscr{E}$.

§9. 概率测度之间的变差距离. 角谷 – 海林格距离和海林格积分. 对测度的绝对连续性和奇异性的应用

1. 概率测度间的变差距离 设 $(\Omega, \mathscr{F})$ 是可测空间, $\mathscr{P} = \{P\}$ 是空间 $(\Omega, \mathscr{F})$ 上的概率测度族.

定义 1 设 P 和 $\widetilde{P}$ 为 $\mathscr{P}$ 中的两个测度. 称变量

$$\operatorname{var}(P - \widetilde{P}) \equiv \sup_{\varphi} \left| \int_{\Omega} \varphi(\omega) \mathrm{d}(P - \widetilde{P}) \right| \tag{1}$$

为 (带符号) 测度 $P - \widetilde{P}$ 的**全变差**, 其中 sup 在一切满足 $|\varphi(\omega)| \leqslant 1$ 的 $\mathscr{F}$ – 可测函数 $\varphi(\omega)$ 类上来求. 称 $\operatorname{var}(P - \widetilde{P})$ 为 P 中测度 P 和 $\widetilde{P}$ 之间的**变差距离**, 记作 $\|P - \widetilde{P}\|$.

引理 1 *变差距离*

$$\|P - \widetilde{P}\| = 2 \sup_{A \in \mathscr{F}} |P(A) - \widetilde{P}(A)|. \tag{2}$$

证明 由于对于任意 $A \in \mathscr{F}$,

$$P(A) - \widetilde{P}(A) = \widetilde{P}(\overline{A}) - P(\overline{A}),$$

则

$$2|P(A)-\widetilde{P}(A)| = |P(A)-\widetilde{P}(A)| + |P(\overline{A})-\widetilde{P}(\overline{A})| \leqslant \|P-\widetilde{P}\|,$$

其中由 (1) 式得最后一个不等式.

为证明相反的不等式, 借助于*带符号测度* $\mu \equiv P-\widetilde{P}$ 的哈恩 (H. Hahn) 分解 (例如, 见 [33, 第六章 §5] 或 [70 第 121 页]). 根据这一分解, 测度 μ 表示为 $\mu=\mu_+-\mu_-$, 其中 μ_+ 和 μ_- (测度 μ 的上变差 μ_+ 和下变差 μ_-) 具有如下形式:

$$\mu_+(A)=\int_{A\cap M}\mathrm{d}\mu, \quad \mu_-(A)=\int_{A\cap \overline{M}}\mathrm{d}\mu, \quad A\in\mathscr{F},$$

其中 M 是 $\mathscr{F}$ 中的集合. 这时

$$\mathrm{var}\mu=\mathrm{var}\mu_+ + \mathrm{var}\mu_- = \mu_+(\Omega)+\mu_-(\Omega).$$

由于

$$\mu_+(\Omega)=P(M)-\widetilde{P}(M), \quad \mu_-(\Omega)=\widetilde{P}(\overline{M})-P(\overline{M}),$$

可见

$$\|P-\widetilde{P}\| = [P(M)-\widetilde{P}(M)] + [\widetilde{P}(\overline{M})-P(\overline{M})] \leqslant 2\sup_{A\in\mathscr{F}}|P(A)-\widetilde{P}(A)|. \qquad \square$$

定义 2 称概率测度序列 $\{P_n\}_{n\geqslant1}$ *按变差收敛*于测度 P, 记为 $P_n \xrightarrow{var} P$, 如果

$$\|P_n-P\|\to 0, \quad n\to\infty. \tag{3}$$

由这一定义和 §1 的定理 1, 不难看出, 如果定义在度量空间 $(\Omega,\mathscr{F},\rho)$ 上的概率测度按变差收敛, 则必定也弱收敛.

分布按变差的接近程度, 看来是概率分布接近程度最强的形式. 因为, 若两个分布的变差接近, 则实际上在具体的情形下可以认为它们是无区别的. 因此, 可能造成一种印象, 似乎研究变差距离没有太大的概率意义. 然而, 例如在泊松定理中 (第一章 §6), 二项分布向泊松分布的收敛, 就是按变差收敛于 0 (下面 §12 将给出这个距离的上估计.)

下面将要举出数理统计中用到两个测度 P 和 $\widetilde{P}$ 间*变差距离*的例子. 这种情形自然地出现在根据观测结果区分两个统计假设 H 和 $\widetilde{H}$ 的问题中. 考虑两个假设, H: 真实的分布为 P,$\widetilde{H}$: 真实的分布为 $\widetilde{P}$; 要求判断两个概率模型: $(\Omega,\mathscr{F},\mathbf{P})$ 或 $(\Omega,\mathscr{F},\widetilde{\mathbf{P}})$, 哪一个更符合对观测结果的 "统计". 如果把 $\omega\in\Omega$ 解释为观测结果, 则认为取值于 [0,1] 的任意 $\mathscr{F}$– 可测函数 $\varphi=\varphi(\omega)$ 是 (区分假设 H 和 $\widetilde{H}$ 的) *准则*. 准则的统计意义是, $\varphi(\omega)$ 是“当观测结果为 ω 时, 接受假设 $\widetilde{H}$ 的概率”.

我们用*第一类错误概率*和*第二类错误概率*表征区分假设 H 和 $\widetilde{H}$ 的准则的质量:
$\alpha(\varphi)=E\varphi(\omega)$ (=“当 H 真实时, 接受 $\widetilde{H}$ 的概率”),

$\beta(\varphi) = \widetilde{E}[1 - \varphi(\omega)]$ (="当 $\widetilde{H}$ 真实时, 接受 H 的概率"),

其中 E 和 $\widetilde{E}$ 表示按测度 P 和 $\widetilde{P}$ 求平均. 假如假设 H 和 $\widetilde{H}$ 是对等的, 自然认为使误差概率之和 $\alpha(\varphi) + \beta(\varphi)$ 最小的准则 $\varphi^* = \varphi^*(\omega)$ 是**最优的** (只要这样的准则存在).

设

$$\mathscr{E}_r(P, \widetilde{P}) = \inf_{\varphi}[\alpha(\varphi) + \beta(\varphi)]. \tag{4}$$

记

$$Q = \frac{1}{2}(P + \widetilde{P}) \text{ 和 } z = \frac{dP}{dQ},\ \widetilde{z} = \frac{d\widetilde{P}}{dQ},$$

则

$$\mathscr{E}_r(P, \widetilde{P}) = \inf_{\varphi}[E\varphi + \widetilde{E}(1 - \varphi)] = \inf_{\varphi} E_Q[z\varphi + \widetilde{z}(1 - \varphi)] = 1 + \inf_{\varphi} E_Q[\varphi(z - \widetilde{z})]$$

(这里及以后用 E_Q 表示对测度 Q 的数学期望).

不难看到, 在函数

$$\varphi^*(\omega) = I\{\widetilde{z} < z\}$$

上达到 inf, 且因为 $E_Q(z - \widetilde{z}) = 0$, 所以

$$\mathscr{E}_r(P, \widetilde{P}) = 1 - \frac{1}{2}E_Q|z - \widetilde{z}| = 1 - \frac{1}{2}\|P - \widetilde{P}\|, \tag{5}$$

其中由下面的引理 2, 可得最后一个等式. 于是, 由 (5) 式可见, 函数 $\mathscr{E}_r(P, \widetilde{P})$ 表征区分假设的质量, 而 $\mathscr{E}_r(P, \widetilde{P})$ 依赖于由*变差距离*表征的测度 P 和 $\widetilde{P}$ 的接近程度.

引理 2 *设 Q 是某一 σ– 有限测度, 满足 $P \ll Q, \widetilde{P} \ll Q$, 而测度 P 和 $\widetilde{P}$ 关于 Q 的拉东 – 尼科迪姆导数记为*

$$z = \frac{dP}{dQ},\quad \widetilde{z} = \frac{d\widetilde{P}}{dQ},$$

那么,

$$\|P - \widetilde{P}\| = E_Q|z - \widetilde{z}|, \tag{6}$$

而且, 如果 $Q = (P + \widetilde{P})/2$, 则

$$\|P - \widetilde{P}\| = E_Q|z - \widetilde{z}| = 2E_Q|1 - z| = 2E_Q|1 - \widetilde{z}|. \tag{7}$$

证明 对于一切 $\mathscr{F}$– 可测函数 $\psi = \psi(\omega), |\psi(\omega)| \leqslant 1$, 根据 z 和 $\widetilde{z}$ 的定义, 有

$$|E\psi - \widetilde{E}\psi| = |E_Q\psi(z - \widetilde{z})| \leqslant E_Q|\psi||(z - \widetilde{z})| \leqslant E_Q|(z - \widetilde{z})|. \tag{8}$$

因此

$$\|P - \widetilde{P}\| \leqslant E_Q|z - \widetilde{z}|, \tag{9}$$

但是, 对于函数

$$\psi = \operatorname{sign}(\widetilde{z} - z) = \begin{cases} 1, & \widetilde{z} \geqslant z, \\ -1, & \widetilde{z} < z, \end{cases}$$

有

$$|E\psi - \widetilde{E}\psi| = E_Q|(z - \widetilde{z})|. \tag{10}$$

由 (9) 和 (10) 式得所求的 (6) 式. 由于 $z + \widetilde{z} = 2(Q - \text{a.c.})$, 由 (6) 式得 (7) 式. □

系 1 *设 P 和 $\widetilde{P}$ 是 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的两个概率分布, 其 (关于勒贝格测度 $\mathrm{d}x$ 的) 概率密度为 $p(x)$ 和 $\widetilde{p}(x), x \in \mathbb{R}$, 则*

$$\|P - \widetilde{P}\| = \int_{-\infty}^{\infty} |p(x) - \widetilde{p}(x)| \mathrm{d}x. \tag{11}$$

(作为 Q 应取 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的勒贝格测度.)

系 2 *设 P 和 $\widetilde{P}$ 是两个集中在可数个点 $x_1, x_2, \cdots$ 上的离散测度: $P = \{p_1, p_2, \cdots\}, \widetilde{P} = \{\widetilde{p}_1, \widetilde{p}_2, \cdots\}$, 则*

$$\|P - \widetilde{P}\| = \sum_{i=1}^{\infty} |p_i - \widetilde{p}_i|. \tag{12}$$

(作为 Q 应取读数测度, 即 $Q(\{x_i\}) = 1, i = 1, 2, \cdots$)

2. 测度间的角谷 – 海林格 (S. Kakutani–E. Helinger) 距离 我们现在再考虑一种概率测度接近程度的度量, 它在很大程度上与变差测度的接近程度是同种的.

设 P 和 $\widetilde{P}$ 是 $(\Omega, \mathscr{F})$ 上的两个概率测度, 而 Q 是控制 P 和 $\widetilde{P}$ 的另一个概率测度[①], 即 $P \ll Q, \widetilde{P} \ll Q$. 仍记

$$z = \frac{\mathrm{d}P}{\mathrm{d}Q}, \quad \widetilde{z} = \frac{\mathrm{d}\widetilde{P}}{\mathrm{d}Q}.$$

定义 3 称非负量 $\rho(P, \widetilde{P})$ 为测度 P 和 $\widetilde{P}$ 之间的*角谷 – 海林格距离*, 如果

$$\rho^2(P, \widetilde{P}) = \frac{1}{2} E_Q(\sqrt{z} - \sqrt{\widetilde{z}})^2, \tag{13}$$

由于

$$E_Q(\sqrt{z} - \sqrt{\widetilde{z}})^2 = \int_{\Omega} \left(\sqrt{\frac{\mathrm{d}P}{\mathrm{d}Q}} - \sqrt{\frac{\mathrm{d}\widetilde{P}}{\mathrm{d}Q}} \right)^2 \mathrm{d}Q, \tag{14}$$

① 一般称 Q (相对 P 和 $\widetilde{P}$) 为优测度或强测度. —— 译者

则量 $\rho^2(P,\widetilde{P})$ 的书写形式

$$\rho^2(P,\widetilde{P}) = \frac{1}{2}\int_\Omega \left(\sqrt{\mathrm{d}P} - \sqrt{\mathrm{d}\widetilde{P}}\right)^2 \tag{15}$$

就变得清晰明了.

如果设

$$H(P,\widetilde{P}) = E_Q\sqrt{z\widetilde{z}}, \tag{16}$$

则仿照 (15) 式, 可以形式地写为

$$H(P,\widetilde{P}) = \int_\Omega \sqrt{\mathrm{d}P\mathrm{d}\widetilde{P}}. \tag{17}$$

由 (13) 和 (16) 式, 以及由 (15) 和 (17) 式, 可见

$$\rho^2(P,\widetilde{P}) = 1 - H(P,\widetilde{P}). \tag{18}$$

量 $H(P,\widetilde{P})$ 称做 1/2 阶海林格积分. 对于许多应用, 考虑 $\alpha \in (0,1)$ 阶海林格积分 $H(\alpha;P,\widetilde{P})$ 是有益的, 它由下面的公式定义:

$$H(\alpha;P,\widetilde{P}) = E_Q z^\alpha \widetilde{z}^{1-\alpha}, \tag{19}$$

或形式地写为

$$H(\alpha;P,\widetilde{P}) = \int_\Omega (\mathrm{d}P)^\alpha (\mathrm{d}\widetilde{P})^{1-\alpha}. \tag{20}$$

显然 $H(1/2;P,\widetilde{P}) = H(P,\widetilde{P})$.

为保障定义 3 是适定的, 需要证明量 $\rho^2(P,\widetilde{P})$ 与控制测度的选择无关, 且 $\rho(P,\widetilde{P})$ 事实上满足对 "距离" 的要求.

引理 3 1. $\alpha \in (0,1)$ 阶海林格积分 $H(\alpha;P,\widetilde{P})$ (从而 $\rho(P,\widetilde{P})$) 与控制测度 Q 的选择无关.

2. 由 (13) 式定义的函数 ρ, 在概率测度的集合上是度量.

证明 1. 假如 Q' 控制 P 和 $\widetilde{P}$ (即 $P \ll Q', \widetilde{P} \ll Q'$), 则 Q' 也控制测度 $Q = (P+\widetilde{P})/2$. 因此, 只需证明, 若 $Q \ll Q'$, 则

$$E_Q(z^\alpha \widetilde{z}^{1-\alpha}) = E_{Q'}(z')^\alpha (\widetilde{z}')^{1-\alpha},$$

其中

$$z' = \frac{\mathrm{d}P}{\mathrm{d}Q'}, \quad \widetilde{z}' = \frac{\mathrm{d}\widetilde{P}}{\mathrm{d}Q'}.$$

记 $v = \mathrm{d}Q/\mathrm{d}Q'$. 那么 $z' = zv, \widetilde{z}' = \widetilde{z}v$, 且

$$E_Q(z^\alpha \widetilde{z}^{1-\alpha}) = E_{Q'}(vz^\alpha \widetilde{z}^{1-\alpha}) = E_{Q'}(z')^\alpha (\widetilde{z}')^{1-\alpha},$$

于是, 第一个命题得证.

2. 如果 $\rho(P,\widetilde{P})=0$, 则 $z=\widetilde{z}(Q-\text{a. c.})$, 因此 $P=\widetilde{P}$. 显然, 由对称性可见 $\rho(P,\widetilde{P})=\rho(\widetilde{P},P)$. 最后, 设 P,P',P'' 是 3 个测度, 满足 $P\ll Q, P'\ll Q, P''\ll Q$, 且

$$z=\frac{\mathrm{d}P}{\mathrm{d}Q},\quad z'=\frac{\mathrm{d}P'}{\mathrm{d}Q},\quad z''=\frac{\mathrm{d}P''}{\mathrm{d}Q}.$$

利用对于 $L^2(\Omega,\mathscr{F},Q)$ 中的范数成立三角形不等式, 得

$$[E_Q(\sqrt{z}-\sqrt{z''})^2]^{1/2}\leqslant[E_Q(\sqrt{z}-\sqrt{z'})^2]^{1/2}+[E_Q(\sqrt{z'}-\sqrt{z''})^2]^{1/2},$$

即

$$\rho(P,P'')\leqslant\rho(P,P')+\rho(P',P'').$$ □

由定义 (19) 和傅比尼定理 (第二章 §6) 可以直接推出, 当测度 P 和 $\widetilde{P}$ 是测度的直积时 (第二章 §6 第 10 小节): $P=P_1\times\cdots\times P_n, \widetilde{P}=\widetilde{P}_1\times\cdots\times\widetilde{P}_n$, 则测度 P 和 $\widetilde{P}$ 间的海林格积分等于相应积分的积:

$$H(\alpha;P,\widetilde{P})=\prod_{i=1}^{n}H(\alpha;P_i,\widetilde{P}_i).$$

下面的定理揭示了变差距离与角谷 – 海林格距离 (或等价地, 海林格积分) 间的联系. 特别, 它表明这些距离决定概率测度空间在 $(\Omega,\mathscr{F})$ 上的同一拓扑.

定理 1 下列不等式成立:

$$2[1-H(P,\widetilde{P})]\leqslant\|P-\widetilde{P}\|\leqslant\sqrt{8[1-H(P,\widetilde{P})]},\tag{21}$$

$$\|P-\widetilde{P}\|\leqslant2\sqrt{1-H^2(P,\widetilde{P})}.\tag{22}$$

特别,

$$2\rho^2(P,\widetilde{P})\leqslant\|P-\widetilde{P}\|\leqslant\sqrt{8}\rho(P,\widetilde{P}).\tag{23}$$

证明 由于 $H(P,\widetilde{P})\leqslant1$, 而对于 $0\leqslant x\leqslant1, 1-x^2\leqslant2(1-x)$, 则由 (22) 式得 (21) 式的第二个不等式, 而 (22) 式由下面的一系列不等式 (其中 $Q=(P+\widetilde{P})/2$) 可得:

$$\begin{aligned}\frac{1}{2}\|P-\widetilde{P}\|&=E_Q|1-z|\leqslant\sqrt{E_Q|1-z|^2}=\sqrt{1-E_Qz(2-z)}=\sqrt{1-E_Qz\widetilde{z}}\\&=\sqrt{1-E_Q\big(\sqrt{z\widetilde{z}}\big)^2}\leqslant\sqrt{1-\big(E_Q\sqrt{z\widetilde{z}}\big)^2}=\sqrt{1-H^2(P,\widetilde{P})}.\end{aligned}$$

最后, 由不等式

$$\frac{1}{2}[\sqrt{z}-\sqrt{2-z}]^2\leqslant|z-1|,\ z\in[0,2],$$

有 (其中 $Q=(P+\widetilde{P})/2$),

$$1-H(P,\widetilde{P})=\rho^2(P,\widetilde{P})=\frac{1}{2}E_Q[\sqrt{z}-\sqrt{2-z}]^2\leqslant E_Q|z-1|=\frac{1}{2}\|P-\widetilde{P}\|.$$

由此可得 (21) 式的第一个不等式. □

注 类似地可以证明, 对于任意 $\alpha\in(0,1)$, 有

$$2[1-H(\alpha;P,\widetilde{P})]\leqslant\|P-\widetilde{P}\|\leqslant\sqrt{c_\alpha[1-H(\alpha;P,\widetilde{P})]},\tag{24}$$

其中 c_α 是某一常数.

系 3 设 P 和 $P^n, n\geqslant 1$, 是 $(\Omega,\mathscr{F})$ 上的概率测度. 那么, 当 $n\to\infty$ 时, 有

$$\|P^n-P\|\to 0\Leftrightarrow H(P^n,P)\to 1\Leftrightarrow \rho(P^n,P)\to 0,$$
$$\|P^n-P\|\to 2\Leftrightarrow H(P^n,P)\to 0\Leftrightarrow \rho(P^n,P)\to 1.$$

系 4 由于根据 (5)

$$\mathscr{E}_r(P,\widetilde{P})=1-\frac{1}{2}\|P-\widetilde{P}\|,$$

则由 (21) 式和 (22) 式, 有

$$\frac{1}{2}[H^2(P,\widetilde{P})]\leqslant 1-\sqrt{1-H^2(P,\widetilde{P})}\leqslant\mathscr{E}_r(P,\widetilde{P})\leqslant H(P,\widetilde{P}).\tag{25}$$

特别, 设

$$P^n=\underbrace{P\times\cdots\times P}_{n},\quad \widetilde{P}^n=\underbrace{\widetilde{P}\times\cdots\times\widetilde{P}}_{n}$$

相应为 n 个同样测度的直积. 那么, 由于

$$H(P^n,\widetilde{P}^n)=[H(P,\widetilde{P})]^n=\mathrm{e}^{-\lambda n},\quad \lambda=-\ln H(P,\widetilde{P})\geqslant\rho^2(P,\widetilde{P}).$$

则由 (25) 式, 得

$$\frac{1}{2}\mathrm{e}^{-2\lambda n}\leqslant\mathscr{E}_r(P^n,\widetilde{P}^n)\leqslant\mathrm{e}^{-\lambda n}\leqslant\mathrm{e}^{-n\rho^2(P,\widetilde{P})}.\tag{26}$$

用于上面考虑的区分两个统计假设的问题, 由这些不等式可得如下结果.

设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机元, 其概率分布为 P (假设 H) 或 $\widetilde{P}$ (假设 $\widetilde{H}$), 且 $P\neq\widetilde{P}$, 从而 $\rho^2(P,\widetilde{P})>0$. 那么, 表征根据观测结果 $\xi_1,\cdots,\xi_n$, 最优区分假设 H 和 $\widetilde{H}$ 质量的函数 $\mathscr{E}_r(P^n,\widetilde{P}^n)$, 当 $n\to\infty$ 时以指数的速度递降为 0.

3. 海林格积分对测度的绝对连续性和奇异性的应用　利用上面引进的 α 阶海林格积分, 容易表述概率测度绝对连续性和奇异性的条件.

设 P 和 $\widetilde{P}$ 是可测空间 $(\Omega,\mathscr{F})$ 上的两个概率测度. 回忆, 测度 $\widetilde{P}$ 关于测度 P 绝对连续 (记作 $\widetilde{P}\ll P$), 如果对于 $A\in\mathscr{F}$, 每当 $P(A)=0$ 时必有 $\widetilde{P}(A)=0$. 如果 $\widetilde{P}\ll P$ 且 $P\ll\widetilde{P}$, 则称 P 和 $\widetilde{P}$ 等价 $(\widetilde{P}\sim P)$. 测度 P 和 $\widetilde{P}$ 称做奇异的或正交的 $(\widetilde{P}\perp P)$, 如果存在 $A\in\mathscr{F}$, 使 $P(A)=1$ 和 $\widetilde{P}(\overline{A})=1$ (即测度 P 和 $\widetilde{P}$ "处于" 不同的集合中).

设 Q 是概率测度, 而 $P\ll Q,\widetilde{P}\ll Q$,

$$z=\frac{\mathrm{d}P}{\mathrm{d}Q},\quad \widetilde{z}=\frac{\mathrm{d}\widetilde{P}}{\mathrm{d}Q}.$$

定理 2　下列条件等价:

(a) $\widetilde{P}\ll P$,

(b) $\widetilde{P}\{z>0\}=1$,

(c) $H(\alpha;P,\widetilde{P})\to 1,\alpha\downarrow 0$.

定理 3　下列条件等价:

(a) $\widetilde{P}\perp P$,

(b) $\widetilde{P}\{z>0\}=0$,

(c) $H(\alpha;P,\widetilde{P})\to 0,\alpha\downarrow 0$,

(d) $H(\alpha;P,\widetilde{P})=0$, 对于一切 $\alpha\in(0,1)$,

(e) $H(\alpha;P,\widetilde{P})=0$, 对于某个 $\alpha\in(0,1)$.

证明　两个定理的证明同时进行. 根据 z 和 $\widetilde{z}$ 的定义

$$P\{z=0\}=E_Q[zI(z=0)]=0, \tag{27}$$

$$\begin{aligned}\widetilde{P}(A\cap\{z>0\})&=E_Q[\widetilde{z}I(A\cap\{z>0\})]=E_Q\left[\widetilde{z}\frac{z}{z}I(A\cap\{z>0\})\right]\\&=E\left[\frac{\widetilde{z}}{z}I(A\cap\{z>0\})\right]=E\left[\frac{\widetilde{z}}{z}I(A)\right].\end{aligned} \tag{28}$$

从而, "**勒贝格分解**" 成立:

$$\widetilde{P}(A)=E\left[\frac{\widetilde{z}}{z}I(A)\right]+\widetilde{P}(A\cap\{z=0\}),\quad A\in\mathscr{F}, \tag{29}$$

其中随机变量 $Z=\widetilde{z}/z$ 称做测度 $\widetilde{P}$ (的 "绝对连续分量") 关于测度 P 的勒贝格导数, 并且记作 $\mathrm{d}\widetilde{P}/\mathrm{d}P$ (对照第二章 §6 拉东 - 尼科迪姆定理的注).

由此立即得出两个定理中 (a) 和 (b) 的等价性.

此外, 由于

$$z^{\alpha}\widetilde{z}^{1-\alpha}\to\widetilde{z}I(z>0),\quad \alpha\downarrow 0,$$

而对于 $\alpha \in (0,1)$

$$0 \leqslant z^{\alpha}\widetilde{z}^{1-\alpha} \leqslant \alpha z + (1-\alpha)\widetilde{z} \leqslant z + \widetilde{z},$$

且 $E_Q(z+\widetilde{z}) = 2$, 故根据勒贝格控制收敛定理

$$\lim_{\alpha\downarrow 0} H(\alpha; P, \widetilde{P}) = E_Q \widetilde{z} I(z>0) = \widetilde{P}(z>0),$$

而这说明两个定理中 (b) $\Leftrightarrow$ (c).

最后证明, 在第二个定理中 (c) $\Leftrightarrow$ (d) $\Leftrightarrow$ (e). 为此只需注意到

$$H(\alpha; P, \widetilde{P}) = \widetilde{E}\left(\frac{z}{\widetilde{z}}\right)^{\alpha} I(\widetilde{z}>0), \quad \widetilde{P}(\widetilde{z}>0) = 1.$$

这说明, 对于每一个 $\alpha \in (0,1), \widetilde{P}\{z>0\} = 0 \Leftrightarrow H(\alpha; P, \widetilde{P}) = 0$, 由此可得蕴涵关系 (c) $\Leftrightarrow$ (d) $\Leftrightarrow$ (e). □

例 1 设 $P = P_1 \times P_2 \times \cdots, \widetilde{P} = \widetilde{P}_1 \times \widetilde{P}_2 \times \cdots$, 其中 P_k 和 $\widetilde{P}_k$ 是 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的高斯测度, 其密度为

$$p_k(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-a_k)^2}{2}}, \quad \widetilde{p}_k(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\widetilde{a}_k)^2}{2}}.$$

由于通过简单的计算, 得

$$H(\alpha; P, \widetilde{P}) = \prod_{k=1}^{\infty} H(\alpha; P_k, \widetilde{P}_k),$$

其中

$$H(\alpha; P_k, \widetilde{P}_k) = \int_{-\infty}^{\infty} p_k^{\alpha}(x)\widetilde{p}_k^{1-\alpha}(x)\mathrm{d}x = \mathrm{e}^{-\frac{\alpha(1-\alpha)}{2}(a_k-\widetilde{a}_k)^2},$$

得

$$H(\alpha; P, \widetilde{P}) = \mathrm{e}^{-\frac{\alpha(1-\alpha)}{2}\sum\limits_{k=1}^{\infty}(a_k-\widetilde{a}_k)^2}.$$

由定理 2 和 3 得

$$\widetilde{P} \ll P \Leftrightarrow P \ll \widetilde{P} \Leftrightarrow \widetilde{P} \sim P \Leftrightarrow \sum_{k=1}^{\infty}(a_k - \widetilde{a}_k)^2 < \infty,$$

$$\widetilde{P} \perp P \Leftrightarrow \sum_{k=1}^{\infty}(a_k - \widetilde{a}_k)^2 = \infty.$$

例 2 仍设 $P = P_1 \times P_2 \times \cdots, \widetilde{P} = \widetilde{P}_1 \times \widetilde{P}_2 \times \cdots$, 其中 P_k 和 $\widetilde{P}_k$ 是参数相应为 $\lambda_k > 0$ 和 $\widetilde{\lambda}_k > 0$ 的泊松分布. 那么, 不难证明

$$\begin{aligned} &\widetilde{P} \ll P \Leftrightarrow P \ll \widetilde{P} \Leftrightarrow \widetilde{P} \sim P \Leftrightarrow \sum_{k=1}^{\infty}\left(\sqrt{\lambda_k} - \sqrt{\widetilde{\lambda}_k}\right)^2 < \infty, \\ &\widetilde{P} \perp P \Leftrightarrow \sum_{k=1}^{\infty}\left(\sqrt{\lambda_k} - \sqrt{\widetilde{\lambda}_k}\right)^2 = \infty. \end{aligned} \tag{30}$$

4. 练习题

1. 在引理 2 的记号下, 设

$$P \wedge \widetilde{P} = E_Q(z \wedge \widetilde{z}),$$

其中 $z \wedge \widetilde{z} = \min(z, \widetilde{z})$. 证明

$$\|P - \widetilde{P}\| = 2(1 - P \wedge \widetilde{P})$$

(从而,$\mathscr{E}_r(P, \widetilde{P}) = P \wedge \widetilde{P}$).

2. 设 $P, P_n, n \geqslant 1$, 是 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的两个概率测度, 其 (关于勒贝格测度的) 密度为 $p(x), p_n(x), n \geqslant 1$. 假设对关于勒贝格测度几乎所有 $x, p_n(x) \to p(x)$, 证明

$$\|P - P_n\| = \int_{-\infty}^{\infty} |p(x) - p_n(x)| \mathrm{d}x \to 0, \quad n \to \infty$$

(对照第二章 §6 练习题 17).

3. 设 P 和 $\widetilde{P}$ 是两个概率测度. 定义库尔贝克 (S. Kullback) 信息量 —— 相对 $\widetilde{P}$ 有利于 P 的信息量 —— $K(P, \widetilde{P})$, 定义为:

$$K(P, \widetilde{P}) = \begin{cases} E \ln \dfrac{\mathrm{d}P}{\mathrm{d}\widetilde{P}}, & \text{若 } P \ll \widetilde{P}, \\ \infty, & \text{若不然}. \end{cases}$$

证明

$$K(P, \widetilde{P}) \geqslant -2\ln[1 - \rho^2(P, \widetilde{P})] \geqslant 2\rho^2(P, \widetilde{P}).$$

4. 证明公式 (11),(12).

5. 证明不等式 (24).

6. 设 $P, \widetilde{P}, Q$ 是 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的 3 个概率测度, 而 $P * Q$ 和 $\widetilde{P} * Q$ 是它们的卷积 (第二章 §8 第 4 小节). 证明

$$\|P * Q - \widetilde{P} * Q\| \leqslant \|P - \widetilde{P}\|.$$

7. 证明 (30) 式.

8. 设 ξ 和 η 是在可测空间 $(E, \mathscr{E})$ 取值的, 概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的随机元, 证明

$$|\mathbf{P}\{\xi \in A\} - \mathbf{P}\{\eta \in A\}| \leqslant \mathbf{P}\{\xi \neq \eta\}, \quad A \in \mathscr{E}.$$

§10. 概率测度的临近性和完全渐近可区分性

1. 概率测度的临近性和完全渐近可区分性的概念 这些概念, 在数理统计的渐近理论中起重要作用, 对于两个测度序列的情形, 是两个测度绝对连续性和奇异性概念的自然推广.

我们从定义开始.

设 $(\Omega^n, \mathscr{F}^n)_{n\geqslant 1}$ 是某一可测空间序列, 而 $(P^n)_{n\geqslant 1}, (\widetilde{P}^n)_{n\geqslant 1}$ 是概率测度序列, 其中 P^n 和 $\widetilde{P}^n$ 定义在 $(\Omega^n, \mathscr{F}^n)_{n\geqslant 1}$ 上.

定义 1 称测度序列 $(\widetilde{P}^n)$ *临近*序列 (P^n), 记作 $(\widetilde{P}^n) \lhd (P^n)$, 如果对于满足 $P^n(A^n) \to 0, n \to \infty$ 的所有 $A^n \in \mathscr{F}^n$, 有 $\widetilde{P}^n(A^n) \to 0, n \to \infty$.

定义 2 称测度序列 $(\widetilde{P}^n)$ 和 (P^n)*完全渐近可区分*, 简称*可区分*, 记作 $(\widetilde{P}^n) \triangle (P^n)$, 如果存在序列 $n_k \uparrow \infty, k \to \infty$ 与这样的集合 $A^{n_k} \in \mathscr{F}^{n_k}$, 使得:

$$P^{n_k}(A^{n_k}) \to 1, \ \widetilde{P}^{n_k}(A^{n_k}) \to 0, \ k \to \infty.$$

立即可以指出, 可分性是一个对称的概念: $(\widetilde{P}^n) \triangle (P^n) \Leftrightarrow (P^n) \triangle (\widetilde{P}^n)$. 临近性不具备这一条性质. 假如 $(\widetilde{P}^n) \lhd (P^n)$ 且 $(\widetilde{P}^n) \rhd (P^n)$, 则记作 $(\widetilde{P}^n) \lhd\rhd (P^n)$, 并且称该测度序列 $(\widetilde{P}^n)$ 与 (P^n) 为*相互临近的*.

需要指出对于下面的情形: 对于一切 $n \geqslant 1, (\Omega^n, \mathscr{F}^n) = (\Omega, \mathscr{F}), P^n = P, \widetilde{P}^n = \widetilde{P}$, 有

$$(\widetilde{P}^n) \lhd (P^n) \Leftrightarrow \widetilde{P} \ll P, \tag{1}$$

$$(\widetilde{P}^n) \lhd\rhd (P^n) \Leftrightarrow \widetilde{P} \sim P, \tag{2}$$

$$(\widetilde{P}^n) \triangle (P^n) \Leftrightarrow \widetilde{P} \perp P. \tag{3}$$

这些性质以及上面给出的定义说明, 为什么临近性和完全渐近可分性, 常解释为序列 $(\widetilde{P}^n)$ 和 (P^n) 的 "渐近绝对连续性" 和 "渐近奇异性".

2. 无限稠密随机变量序列的情形 下面给出的定理 1 和 2 是 §9 的定理 2 和 3 对测度序列情形的自然推广.

设 $(\Omega^n, \mathscr{F}^n)_{n\geqslant 1}$ 是可测空间序列, Q^n 是 $(\Omega^n, \mathscr{F}^n)(n \geqslant 1)$ 上的概率测度, 而 (ξ^n) 是 $(\Omega^n, \mathscr{F}^n)_{n\geqslant 1}$ 上的随机变量 (一般说来, 是广义随机变量; 见第二章 §4).

定义 3 随机变量序列 (ξ^n) 关于测度序列 Q^n 称做*稠密的* (记作 $(\xi^n|Q^n)$ *稠密*), 若

$$\lim_{N\uparrow\infty} \varlimsup_{n} Q^n(|\xi^n| > N) = 0. \tag{4}$$

(对照 §2 中概率测度族的密度的相应定义.)

下面到处将设

$$Q^n = \frac{P^n + \widetilde{P}^n}{2}, \quad z^n = \frac{\mathrm{d}P^n}{\mathrm{d}Q^n}, \quad \widetilde{z}^n = \frac{\mathrm{d}\widetilde{P}^n}{\mathrm{d}Q^n}.$$

此外, 记

$$Z^n = \frac{\widetilde{z}^n}{z^n} \tag{5}$$

为测度 $\widetilde{P}^n$ 关于测度 P^n 的勒贝格导数, 并认为 $2/0=\infty$ (见 §9(29) 式). 注意, 假如 $\widetilde{P}^n \ll P^n$, 则 Z^n 恰好是测度 $\widetilde{P}^n$ 关于测度 P^n 的密度 $\mathrm{d}\widetilde{P}^n/\mathrm{d}P^n$ 的一种变式 (第二章 §6).

对于以后, 有益地指出, 由于

$$P^n\left(z^n \leqslant \frac{1}{N}\right) = E_{Q^n}\left(z^n I\left(z^n \leqslant \frac{1}{N}\right)\right) \leqslant \frac{1}{N}, \tag{6}$$

且 $Z^n \leqslant 2/z^n$, 则

$$\left(\frac{1}{z^n}\middle| P^n\right) \text{ 稠密}, \quad (Z^n|P^n) \text{ 稠密}. \tag{7}$$

定理 1 下列各条件等价:

(a) $(\widetilde{P}^n) \lhd (P^n)$,

(b) $\left(\frac{1}{z^n}\middle|\widetilde{P}^n\right)$ 稠密,

(b′) $(Z^n|\widetilde{P}^n)$ 稠密,

(c) $\lim\limits_{\alpha\downarrow 0}\varliminf\limits_n H(\alpha; P^n, \widetilde{P}^n) = 1$.

定理 2 下列各条件等价:

(a) $(\widetilde{P}^n) \triangle (P^n)$,

(b) 对于任意 $\varepsilon > 0, \varliminf\limits_n \widetilde{P}^n(z^n \geqslant \varepsilon) = 0$,

(b′) 对于任意 $N > 0, \varlimsup\limits_n \widetilde{P}^n(Z^n \leqslant N) = 0$,

(c) $\lim\limits_{\alpha\downarrow 0}\varliminf\limits_n H(\alpha; P^n, \widetilde{P}^n) = 0$,

(d) 对于任意 $\alpha \in (0,1), \varliminf\limits_n H(\alpha; P^n, \widetilde{P}^n) = 0$,

(e) 对于某个 $\alpha \in (0,1), \varliminf\limits_n H(\alpha; P^n, \widetilde{P}^n) = 0$.

定理 1 的证明.

(a)⇒(b). 假如 (b) 不成立, 则存在 $\varepsilon > 0$ 和这样的序列 $n_k \uparrow \infty$, 使 $\widetilde{P}^{n_k}(z^{n_k} < 1/n_k) \geqslant \varepsilon$. 然而, 由 (6) 式, 有 $P^{n_k}(z^{n_k} < 1/n_k) \leqslant 1/n_k \to 0, k \to \infty$, 而这与假设 $(\widetilde{P}^n) \lhd (P^n)$ 矛盾.

(b)⇔(b′). 只需注意到 $Z^n = (2/z^n) - 1$.

(b)⇒(a). 设 $A^n \in \mathscr{F}^n$ 且 $P^n(A^n) \to 0, n \to \infty$, 有

$$\begin{aligned}\widetilde{P}^n(A^n) &\leqslant \widetilde{P}^n(z^n \leqslant \varepsilon) + E_{Q^n}(\widetilde{z}^n I(A^n \cap \{z^n > \varepsilon\})) \\ &\leqslant \widetilde{P}^n(z^n \leqslant \varepsilon) + \frac{2}{\varepsilon} E_{Q^n}(z^n I(A^n)) = \widetilde{P}^n(z^n \leqslant \varepsilon) + \frac{2}{\varepsilon} P^n(A^n).\end{aligned}$$

因此,

$$\varlimsup_n \widetilde{P}^n(A^n) \leqslant \varlimsup_n \widetilde{P}^n(z^n \leqslant \varepsilon), \quad \varepsilon > 0.$$

命题 (b) 等价于

$$\lim_{\varepsilon\downarrow 0}\varlimsup_n \widetilde{P}^n(z^n \leqslant \varepsilon) = 0.$$

于是, $\widetilde{P}^n(A^n)\to 0$,即 (b)⇒(a).

(b)⇒(c). 设 $\varepsilon>0$, 则

$$H(\alpha;P^n,\widetilde{P}^n)=E_{Q^n}[(z^n)^\alpha(\widetilde{z}^n)^{1-\alpha}]\geqslant E_{Q^n}\left[\left(\frac{z^n}{\widetilde{z}^n}\right)^\alpha I(z^n\geqslant\varepsilon)I(\widetilde{z}^n>0)\widetilde{z}^n\right]$$
$$=E_{\widetilde{P}^n}\left[\left(\frac{z^n}{\widetilde{z}^n}\right)^\alpha I(z^n\geqslant\varepsilon)\right]\geqslant\left(\frac{\varepsilon}{2}\right)^\alpha\widetilde{P}^n(z^n\geqslant\varepsilon).\tag{8}$$

由于 $z^n+\widetilde{z}^n=2$, 可见对于 $\varepsilon>0$,

$$\lim_{\alpha\downarrow 0}\varliminf_n H(\alpha;P^n,\widetilde{P}^n)\geqslant\lim_{\alpha\downarrow 0}\left(\frac{\varepsilon}{2}\right)^\alpha\varliminf_n\widetilde{P}(z^n\geqslant\varepsilon)=\varliminf_n\widetilde{P}^n(z^n\geqslant\varepsilon).\tag{9}$$

由于 (b)$\lim\limits_{\varepsilon\downarrow 0}\varliminf\limits_n\widetilde{P}^n(z^n\geqslant\varepsilon)=1$. 因此, 由于 (9) 式和 $H(\alpha;P^n,\widetilde{P}^n)\leqslant 1$, 可得命题 (c).

(c)⇒(b). 设 $\delta\in(0,1)$, 则

$$\begin{aligned}H(\alpha;P^n,\widetilde{P}^n)&=E_{Q^n}[(z^n)^\alpha(\widetilde{z}^n)^{1-\alpha}I(z^n<\varepsilon)]+E_{Q^n}[(z^n)^\alpha(\widetilde{z}^n)^{1-\alpha}I(z^n\geqslant\varepsilon,\widetilde{z}^n\leqslant\delta)]\\&\quad+E_{Q^n}[(z^n)^\alpha(\widetilde{z}^n)^{1-\alpha}I(z^n\geqslant\varepsilon,\widetilde{z}^n>\delta)]\\&\leqslant 2\varepsilon^\alpha+2\delta^{1-\alpha}+E_{Q^n}\left[\widetilde{z}^n\left(\frac{z^n}{\widetilde{z}^n}\right)^\alpha I(z^n\geqslant\varepsilon,\widetilde{z}^n>\delta)\right]\\&\leqslant 2\varepsilon^\alpha+2\delta^{1-\alpha}+\left(\frac{2}{\delta}\right)^\alpha\widetilde{P}^n(z^n\geqslant\varepsilon).\end{aligned}\tag{10}$$

因而, 对于一切 $\alpha\in(0,1),\delta\in(0,1)$, 有

$$\lim_{\varepsilon\downarrow 0}\varliminf_n\widetilde{P}^n(z^n\geqslant\varepsilon)\geqslant\left(\frac{\delta}{2}\right)^\alpha\varliminf_n H(\alpha;P^n,\widetilde{P}^n)-\frac{2\delta}{2^\alpha}.$$

首先令 $\alpha\downarrow 0$ 并利用 (c), 然后令 $\delta\downarrow 0$, 得

$$\lim_{\varepsilon\downarrow 0}\varliminf_n\widetilde{P}^n(z^n\geqslant\varepsilon)\geqslant 1.$$

于是, (b) 的正确性得证. □

定理 2 的证明.

(a)⇒(b). 设 $(\widetilde{P}^n)\triangle(P^n), n_k\uparrow\infty$, 而 $A^{n_k}\in\mathscr{F}^{n_k}$, 则 $P^{n_k}(A^{n_k})\to 1,\widetilde{P}^{n_k}(A^{n_k})\to 0$. 那么, 注意到 $z^n+\widetilde{z}^n=2$, 可见

$$\begin{aligned}\widetilde{P}^{n_k}(z^{n_k}\geqslant\varepsilon)&\leqslant\widetilde{P}^{n_k}(A^{n_k})+E_{Q^{n_k}}\left[z^{n_k}\times\frac{\widetilde{z}^{n_k}}{z^{n_k}}I(\overline{A}^{n_k})I(z^{n_k}\geqslant\varepsilon)\right]\\&=\widetilde{P}^{n_k}(A^{n_k})+M_{P^{n_k}}\left[\frac{\widetilde{z}^{n_k}}{z^{n_k}}I(\overline{A}^{n_k})I(z^{n_k}\geqslant\varepsilon)\right]\\&\leqslant\widetilde{P}^{n_k}(A^{n_k})+\frac{2}{\varepsilon}P^{n_k}(\overline{A}^{n_k}).\end{aligned}$$

从而, $\widetilde{P}^{n_k}(z^{n_k}\geqslant\varepsilon)\to 0$ 故 (b) 成立.

(b)⇒(a). 如果 (b) 成立, 则存在序列 $n_k \uparrow \infty$, 使

$$\widetilde{P}^{n_k}\left(z^{n_k} \geqslant \frac{1}{k}\right) \leqslant \frac{1}{k} \to 0, \quad k \to \infty.$$

由此并注意到 (见 (6) 式)

$$\widetilde{P}^{n_k}\left(z^{n_k} \geqslant \frac{1}{k}\right) \geqslant 1 - \frac{1}{k},$$

得命题 (a).

(b)⇔(b′). 只需注意到 $Z^n = (2/z^n) - 1$.

(b)⇒(d). 由 (10) 式和 (b), 对于属于区间 (0,1) 的任意 ε 和 δ, 有

$$\underline{\lim_n} H(\alpha; P^n, \widetilde{P}^n) \leqslant 2\varepsilon^\alpha + 2\delta^{1-\alpha}.$$

因此,(d) 成立.

(d)⇒(c) 和 (d)⇒(e) 显然.

最后, 由 (8) 式, 有

$$\underline{\lim_n} \widetilde{P}^n(z^n \geqslant \varepsilon) \leqslant \left(\frac{2}{\varepsilon}\right)^\alpha \underline{\lim_n} H(\alpha; P^n, \widetilde{P}^n).$$

由于 $(2/\varepsilon)^\alpha \to 1, \alpha \downarrow 0$, 可见 (c)⇒(b) 和 (e)⇒(b). □

3. 独立观测概型 考虑对应于*独立*观测概型的一种特殊情形. 在这种情形下, 积分 $H(\alpha; P^n, \widetilde{P}^n)$ 的计算和定理 1 和 2 的应用, 都没有大的困难.

假设测度 P^n 和 $\widetilde{P}^n$ 是*测度的直积*:

$$P^n = P_1 \times \cdots \times P_n, \quad \widetilde{P}^n = \widetilde{P}_1 \times \cdots \times \widetilde{P}_n, \quad n \geqslant 1,$$

其中 P_k 和 $\widetilde{P}_k$ 是 $(\Omega_k, \mathscr{F}_k), k \geqslant 1$, 上的测度.

由于在这种情形下,

$$H(\alpha; P^n, \widetilde{P}^n) = \prod_{k=1}^n H(\alpha; P_k, \widetilde{P}_k) = \exp\left\{\sum_{k=1}^n \ln[1 - (1 - H(\alpha; P_k, \widetilde{P}_k))]\right\},$$

则由定理 1 和 2 得如下结果:

$$(\widetilde{P}^n) \lhd (P^n) \Leftrightarrow \lim_{\alpha\downarrow 0} \overline{\lim_n} \sum_{k=1}^n [1 - H(\alpha; P_k, \widetilde{P}_k)] = 0, \tag{11}$$

$$(\widetilde{P}^n) \triangle (P^n) \Leftrightarrow \overline{\lim_n} \sum_{k=1}^n [1 - H(\alpha; P_k, \widetilde{P}_k)] = \infty. \tag{12}$$

例 设 $(\Omega_k, \mathscr{F}_k) = (\mathbb{R}, \mathscr{B}(\mathbb{R})), \alpha_k \in (0,1)$,

$$P_k(\mathrm{d}x) = I_{[0,1]}(x)\mathrm{d}x, \quad \widetilde{P}_k(\mathrm{d}x) = \frac{1}{1-a_k} I_{[a_k,1]}(x)\mathrm{d}x.$$

由于这里 $H(\alpha; P_k, \widetilde{P}_k) = (1-a_k)^\alpha, \alpha \in (0,1)$, 由 (11) 式和 $H(\alpha; P_k, \widetilde{P}_k) = H(1-\alpha; \widetilde{P}_k, P_k)$, 可得

$$(\widetilde{P}^n) \lhd (P^n) \Leftrightarrow \overline{\lim_n} na_n < \infty, \quad \text{即} \quad a_n = O\left(\frac{1}{n}\right),$$

$$(P^n) \lhd (\widetilde{P}^n) \Leftrightarrow \overline{\lim_n} na_n = 0, \quad \text{即} \quad a_n = o\left(\frac{1}{n}\right),$$

$$(\widetilde{P}^n) \vartriangle (P^n) \Leftrightarrow \overline{\lim_n} na_n = \infty.$$

4. 练习题

1. 设 $P^n = P_1^n \times \cdots \times P_n^n, \widetilde{P}^n = \widetilde{P}_1^n \times \cdots \times \widetilde{P}_n^n, n \geqslant 1$, 其中 P_k^n 和 $\widetilde{P}_k^n$ 是参数为 $(a_k^n, 1)$ 和 $(\widetilde{a}_k^n, 1)$ 的高斯测度. 求使 $(\widetilde{P}^n) \lhd (P^n), (\widetilde{P}^n) \vartriangle (P^n)$ 成立, a_k^n 和 $\widetilde{a}_k^n$ 应满足的条件.

2. 设 $P^n = P_1^n \times \cdots \times P_n^n, \widetilde{P}^n = \widetilde{P}_1^n \times \cdots \times \widetilde{P}_n^n, n \geqslant 1$, 其中 P_k^n 和 $\widetilde{P}_k^n$ 是 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的概率测度:$P_k^n(\mathrm{d}x) = I_{[0,1]}(x)\mathrm{d}x, \widetilde{P}_k^n(\mathrm{d}x) = I_{[a_n, 1+a_n]}(x)\mathrm{d}x, 0 \leqslant a_n \leqslant 1$. 证明 $H(\alpha; P_k^n, \widetilde{P}_k^n) = 1 - a_n$ 且

$$(\widetilde{P}^n) \lhd (P^n) \Leftrightarrow (P^n) \lhd (\widetilde{P}^n) \Leftrightarrow \overline{\lim_n} na_n = 0,$$

$$(\widetilde{P}^n) \vartriangle (P^n) \Leftrightarrow \overline{\lim_n} na_n = \infty.$$

3. 设 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 1})$ 是滤波度量空间, 即引进了 $\sigma-$ 代数流 $(\mathscr{F}_n)_{n\geqslant 0}$ 的可测空间 $(\Omega, \mathscr{F})$, 其中 $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}$. 假设 $\mathscr{F} = \sigma(\bigcup_n \mathscr{F}_n)$. 设 P 和 $\widetilde{P}$ 是 $(\Omega, \mathscr{F})$ 上的两个概率测度, 而 $P_n = P|\mathscr{F}_n, \widetilde{P}_n = \widetilde{P}|\mathscr{F}_n$ 是 P 和 $\widetilde{P}$ 在 $\mathscr{F}_n$ 上的收缩. 证明:

$$(P^n) \lhd (P^n) \Leftrightarrow P \ll P,$$

$$(\widetilde{P}^n) \lhd \rhd (P^n) \Leftrightarrow \widetilde{P} \sim P,$$

$$(\widetilde{P}^n) \vartriangle (P^n) \Leftrightarrow \widetilde{P} \perp P.$$

§11. 中心极限定理的收敛速度

1. 中心极限定理中收敛速度的估计 设 $\xi_{n1}, \cdots, \xi_{nn}$ 是独立随机变量序列, $S_n = \xi_{n1} + \cdots + \xi_{nn}, F_n(x) = \mathbf{P}\{S_n \leqslant x\}, n \geqslant 1$. 如果 $S_n \xrightarrow{d} N(0,1)$, 则对于任意 $x \in \mathbb{R}, F_n(x) \to \Phi(x)$. 由于函数 $\Phi(x)$连续, 则这里实际上是一致收敛 (§1 练习题 5):

$$\sup_x |F_n(x) - \Phi(x)| \to 0, \quad n \to \infty. \tag{1}$$

自然提出关于 (1) 式的收敛速度问题. 我们考虑下面的情形:

$$S_n = \frac{\xi_1 + \cdots + \xi_n}{\sigma\sqrt{n}}, \quad n \geqslant 1,$$

其中 $\xi_1, \xi_n, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}\xi_k = 0, \mathbf{D}\xi_k = \sigma^2 > 0$, 而 $\mathbf{E}|\xi_1|^3 < \infty$.

定理 (贝里 – 埃森 [A. C. Berry-C. G. Esseen]) 有如下估计

$$\sup_x |F_n(x) - \Phi(x)| \leqslant \frac{C\mathbf{E}|\xi_1|^3}{\sigma^3\sqrt{n}}, \tag{2}$$

其中 C 是绝对常数 ($(2\pi)^{-1/2} \leqslant C < 0.7655$).

证明 为简单计, 设 $\sigma^2 = 1$ 和 $\beta_3 = \mathbf{E}|\xi_1|^3 < \infty$. 由埃森不等式 (第二章 §12 第 10 小节), 有

$$\sup_x |F_n(x) - \Phi(x)| \leqslant \frac{2}{\pi}\int_0^T \left|\frac{f_n(t) - \varphi(t)}{t}\right| \mathrm{d}t + \frac{24}{\pi T} \times \frac{1}{\sqrt{2\pi}}, \tag{3}$$

其中

$$\varphi(t) = \mathrm{e}^{-\frac{t^2}{2}}, \quad f_n(t) = \left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n,$$

而 $f(t) = \mathbf{E}\mathrm{e}^{it\xi_1}$.

式 (3) 中的正数 T 可以任意选. 设

$$T = \frac{\sqrt{n}}{5\beta_3}.$$

下面将要证明, 对于这样选择的 T, 有

$$|f_n(t) - \varphi(t)| \leqslant \frac{7\beta_3}{6\sqrt{n}}|t|^3\mathrm{e}^{-\frac{t^2}{4}}, \quad |t| \leqslant T. \tag{4}$$

考虑到这一不等式, 由 (3) 式立即得到所要证明的不等式 (2), 而其中 C 是绝对常数. (更加精确的计算表明, C 的值小于 0.7655; 参见 [88 第 5 章 §4.3].)

我们现在证明不等式 (4).

由第二章 §12 的 (18) 式 ($n = 3, \mathbf{E}\xi_1 = 0, \mathbf{E}\xi_1^2 = 1$, 而 $\mathbf{E}|\xi_1|^3 < \infty$), 可见

$$f(t) = \mathbf{E}\mathrm{e}^{it\xi_1} = 1 - \frac{t^2}{2} + \frac{(it)^3}{6}[\mathbf{E}\xi_1^3(\cos\theta_1 t\xi_1 + i\sin\theta_2 t\xi_1)], \tag{5}$$

其中 $|\theta_1| \leqslant 1, |\theta_2| \leqslant 1$. 因此

$$f\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + \frac{(it)^3}{6n^{3/2}}\left[\mathbf{E}\xi_1^3\left(\cos\theta_1\frac{t}{\sqrt{n}}\xi_1 + i\sin\theta_2\frac{t}{\sqrt{n}}\xi_1\right)\right].$$

如果 $|t| \leqslant T = \sqrt{n}/(5\beta_3)$, 则不等式 $\beta_3 \geqslant \sigma^3 = 1$ (见第二章 §6(28) 式), 可见

$$1 - \left|f\left(\frac{t}{\sqrt{n}}\right)\right| \leqslant \left|1 - f\left(\frac{t}{\sqrt{n}}\right)\right| \leqslant \frac{t^2}{2n} + \frac{|t|^3\beta_3}{3n^{3/2}} \leqslant \frac{1}{25}.$$

从而, 当 $|t| \leqslant T$ 时, 有

$$\left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n = \mathrm{e}^{n \ln f\left(\frac{t}{\sqrt{n}}\right)}, \tag{6}$$

其中 $\ln\ z$ 应理解为复数 z 的对数的主值 ($\ln\ z = \ln|z| + i\arg\ z, -\pi < \arg z \leqslant \pi$).

由于 $\beta_3 < \infty$, 则由带拉格朗日余项的泰勒公式 (亦对照第二章 §12 (35) 式), 因为 $s_{\xi_1}^{(1)} = \mathbf{E}\xi_1 = 0, s_{\xi_1}^{(2)} = \sigma^2 = 1$, 可见

$$\begin{aligned}\ln\ f\left(\frac{t}{\sqrt{n}}\right) &= \frac{it}{\sqrt{n}} s_{\xi_1}^{(1)} + \frac{(it)^2}{2n} s_{\xi_1}^{(2)} + \frac{(it)^3}{6n^{3/2}} (\ln\ f)''' \left(\frac{\theta\, t}{\sqrt{n}}\right) \\ &= -\frac{t^2}{2n} + \frac{(it)^3}{6n^{3/2}} (\ln\ f)''' \left(\frac{\theta\, t}{\sqrt{n}}\right),\ |\theta| \leqslant 1.\end{aligned} \tag{7}$$

其次

$$\begin{aligned}[\ln\ f(s)]''' &= \frac{f'''(s)f^2(s) - 3f''(s)f'(s)f(s) + 2[f'(s)]^3}{f^3(s)} \\ &= \frac{\mathbf{E}[(i\xi_1)^3\mathrm{e}^{i\xi_1 s}]f^2(s) - 3\mathbf{E}[(i\xi_1)^2\mathrm{e}^{i\xi_1 s}]\mathbf{E}[(i\xi_1)\mathrm{e}^{i\xi_1 s}]f(s) + 2\mathbf{E}[(i\xi_1)\mathrm{e}^{i\xi_1 s}]^3}{f^3(s)}.\end{aligned}$$

因此, 注意到对于 $|t| \leqslant T, |f(s)| \leqslant 1$, 有 $|f(t/\sqrt{n})| \geqslant 24/25$, 得

$$\left|(\ln\ f)''' \left(\frac{\theta\, t}{\sqrt{n}}\right)\right| \leqslant \frac{\beta_3 + 3\beta_1\beta_2 + 2\beta_1^3}{\left(\frac{24}{25}\right)^3} \leqslant 7\beta_3 \tag{8}$$

($\beta_k = \mathbf{E}|\xi_1|^k, k = 1, 2, 3; \beta_1 \leqslant \beta_2^{1/2} \leqslant \beta_3^{1/3}$; 见第二章 §6 (28) 式).

运用不等式 $|\mathrm{e}^z - 1| \leqslant |z|\mathrm{e}^{|z|}$, 由 (6)~(8) 式, 对于 $|t| \leqslant T = \sqrt{n}/(5\beta_3)$, 得

$$\begin{aligned}\left|\left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n - \mathrm{e}^{-\frac{t^2}{2}}\right| &= \left|\mathrm{e}^{n \ln\ f\left(\frac{t}{\sqrt{n}}\right)} - \mathrm{e}^{-\frac{t^2}{2}}\right| \\ &\leqslant \frac{7\beta_3|t|^3}{6\sqrt{n}} \exp\left\{-\frac{t^2}{2} + \frac{7\beta_3|t|^3}{6\sqrt{n}}\right\} \leqslant \frac{7\beta_3|t|^3}{6\sqrt{n}} \mathrm{e}^{-\frac{t^2}{4}}.\end{aligned}$$ □

注 需要指出, 在缺乏关于可和随机变量的性质的补充假设的情况下, 估计式 (2) 的阶数和估计 $C \geqslant (2\pi)^{-1/2}$ 不能够改进. 事实上, 设 $\xi_1, \xi_n, \cdots$ 是独立同分布伯努利随机变量, 且

$$\mathbf{P}\{\xi_k = +1\} = \mathbf{P}\{\xi_k = -1\} = \frac{1}{2}.$$

由对称性, 可见

$$2\mathbf{P}\left\{\sum_{k=1}^{2n}\xi_k<0\right\}+\mathbf{P}\left\{\sum_{k=1}^{2n}\xi_k=0\right\}=1,$$

因而, 根据斯特林公式 (第一章 §2 (6) 式), 有

$$\left|\mathbf{P}\left\{\sum_{k=1}^{2n}\xi_k<0\right\}-\frac{1}{2}\right|=\frac{1}{2}\mathbf{P}\left\{\sum_{k=1}^{2n}\xi_k=0\right\}=\frac{1}{2}\mathrm{C}_{2n}^{n}2^{-2n}\sim\frac{1}{2\sqrt{\pi n}}=\frac{1}{\sqrt{(2\pi)(2n)}}.$$

特别, 由此可见, 常数 C 不小于 $(2\pi)^{-1/2}$, 且

$$\mathbf{P}\left\{\sum_{k=1}^{2n}\xi_k=0\right\}\sim\frac{1}{\sqrt{\pi n}},\quad n\to\infty. \tag{9}$$

2. 练习题

1. 证明 (8) 式.

2. 设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量, 且 $\mathbf{E}\xi_1=0,\mathbf{D}\xi_1^2=\sigma^2$, 而 $\mathbf{E}|\xi_1|^3<\infty$. 熟知, 对于一切 $-\infty<x<\infty$ 有如下不均匀不等式估计:

$$|F_n(x)-\Phi(x)|\leqslant\frac{C\mathbf{E}|\xi_1|^3}{\sigma^3\sqrt{n}}\times\frac{1}{(1+|x|)^3},$$

证明此不等式, 至少对于伯努利随机变量证明之.

3. 设 $(\xi_k)_{k\geqslant1}$ 是独立同分布随机变量序列, 分别以概率 1/2 取 ±1 为值. 设

$$\varphi_2(t)=\mathbf{E}\mathrm{e}^{it\xi_1}=\frac{1}{2}(\mathrm{e}^{it}+\mathrm{e}^{-it}).$$

仿照拉普拉斯, 证明

$$\mathbf{P}\{S_{2n}=0\}=\frac{1}{\pi}\int_0^{\pi}\varphi_2^n(t)\mathrm{d}t\sim\frac{1}{\sqrt{\pi n}},\quad n\to\infty,$$

其中 $S_k=\xi_1+\cdots+\xi_k$.

4. 设 $(\xi_k)_{k\geqslant0}$ 是独立同分布随机变量序列, 各取 $2a+1$ 个整数值: $0,\pm1,\cdots,\pm a$. 设

$$\varphi_{2a+1}(t)=\mathbf{E}\mathrm{e}^{it\xi_1}=\frac{1}{1+2a}\left(1+2\sum_{k=1}^{a}\cos tk\right).$$

仍仿照拉普拉斯, 证明

$$\mathbf{P}\{S_n=0\}=\frac{1}{\pi}\int_0^{\pi}\varphi_{2a+1}^n(t)\mathrm{d}t\sim\frac{\sqrt{3}}{\sqrt{2\pi(a+1)n}},\quad n\to\infty.$$

特别, 对于 $a=1$, 即对于 ξ_k 取 3 个可能值 $-1,0,1$ 的情形, 证明

$$\mathbf{P}\{S_n=0\}\sim\frac{\sqrt{3}}{2\sqrt{\pi n}},n\to\infty.$$

§12. 泊松定理的收敛速度

1. 泊松定理中收敛速度的估计 设 $\xi_1, \xi_2, \cdots, \xi_n$ 是独立伯努利随机变量, 取 1 和 0 两个可能值, 相应的概率为

$$\mathbf{P}\{\xi_k = 1\} = p_k, \mathbf{P}\{\xi_k = 0\} = q_k (= 1 - p_k), 1 \leqslant k \leqslant n.$$

记 $B = (B_0, B_1, \cdots, B_n)$ 为和 $S = \xi_1 + \cdots + \xi_n$ 的二项分布的概率, 其中 $B_k = \mathbf{P}\{S = k\}$. 再设 $\Pi = (\Pi_0, \Pi_1, \cdots)$ 是参数为 λ 的泊松分布, 其中

$$\Pi_k = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}, k \geqslant 0.$$

如第一章 §6 第 4 小节指出的, 若

$$p_1 = \cdots = p_n, \lambda = np, \tag{1}$$

则对于 B 和 Π 间的变差距离有如下估计 (普罗霍罗夫):

$$\|B - \Pi\| = \sum_{k=0}^{\infty} |B_k - \Pi_k| \leqslant C_1(\lambda) p = C_1(\lambda) \times \frac{\lambda}{n}, \tag{2}$$

其中 $B_{n+1} = B_{n+2} = \cdots = 0$, 而

$$C_1(\lambda) = 2 \min(2, \lambda). \tag{3}$$

对于 $\sum_{k=1}^n p_k = \lambda$ (但 p_k 未必相等) 的情形，卡姆 (L. Le Cam) 证明, 得

$$\|B - \Pi\| = \sum_{k=0}^{\infty} |B_k - \Pi_k| \leqslant C_2(\lambda) \max_{1 \leqslant k \leqslant n} p_k, \tag{4}$$

其中

$$C_2(\lambda) = 2 \min(9, \lambda). \tag{5}$$

由下面将证明的定理可得估计

$$\|B - \Pi\| \leqslant C_3(\lambda) \max_{1 \leqslant k \leqslant n} p_k, \tag{6}$$

其中

$$C_3(\lambda) = 2\lambda. \tag{7}$$

虽然当 $\lambda > 9$ 时 $C_2(\lambda) < C_3(\lambda)$, 即估计 (6) 比估计 (4) 差, 但是我们还是证明 (6) 式, 因为估计 (6) 本质上是初等的, 然而要得到估计 (4) 中 “较好的” 常数, 证明的技术将复杂得多.

2. 估计式 (6) 的证明

定理 设 $\lambda=\sum_{k=1}^{n} p_k$, 则

$$\|B-\Pi\|=\sum_{k=0}^{\infty}|B_k-\Pi_k|\leqslant 2\sum_{k=1}^{n}p_k^2. \tag{8}$$

证明 利用 B 和 Π 中每一个分布是相应分布的卷积:

$$\begin{aligned}B&=B(p_1)*B(p_2)*\cdots*B(p_n),\\ \Pi&=\Pi(p_1)*\Pi(p_2)*\cdots*\Pi(p_n),\end{aligned} \tag{9}$$

理解为相应分布函数的卷积 (见第二章 §8 第 4 小节), 其中 $B(p_k)=(1-p_k,p_k)$ 是参数为 p_k 的在点 0 和 1 的伯努利分布, 而 $\Pi(p_k)$ 是集中在点 $0,1,\cdots$ 参数为 p_k 的泊松分布.

不难证明, 差 $B-\Pi$ 可以表示为:

$$B-\Pi=R_1+\cdots+R_n, \tag{10}$$

其中

$$R_k=[B(p_k)-\Pi_k(p_k)]*F_k, \tag{11}$$

而

$$\begin{aligned}F_1&=\Pi(p_2)*\cdots*\Pi(p_n),\\ F_k&=B(p_1)*\cdots*B(p_{k-1})*\Pi(p_{k+1})*\cdots*\Pi(p_n), 2\leqslant k\leqslant n-1,\\ F_n&=B(p_1)*\cdots*B(p_{n-1}).\end{aligned}$$

由于 §9 的练习题 6, 可见 $\|R_k\|\leqslant\|B(p_k)-\Pi(p_k)\|$. 因此, 由 (10) 式, 立即得

$$\|B-\Pi\|\leqslant\|B(p_k)-\Pi(p_k)\|. \tag{12}$$

注意到 §9 的 (12) 式, 可见计算变差 $\|B(p_k)-\Pi(p_k)\|$ 并不困难:

$$\begin{aligned}\|B(p_k)-\Pi(p_k)\|&=|(1-p_k)-\mathrm{e}^{-p_k}|+|p_k-p_k\mathrm{e}^{-p_k}|+\sum_{j\geqslant 2}\frac{p_k^j\mathrm{e}^{-p_k}}{j!}\\ &=|(1-p_k)-\mathrm{e}^{-p_k}|+|p_k-p_k\mathrm{e}^{-p_k}|+1-\mathrm{e}^{-p_k}-p_k\mathrm{e}^{-p_k}\\ &=2p_k(1-\mathrm{e}^{-p_k})\leqslant 2p_k^2.\end{aligned}$$

于是, 由 (12) 式, 得所要求证明的不等式 (8). □

系 由于

$$\sum_{k=1}^{n}p_k^2\leqslant\lambda\max_{1\leqslant k\leqslant n}p_k,$$

可见估计式成立.

3. 练习题

1. 证明当 $\lambda_k = -\ln(1-p_k)$ 时

$$\|B(p_k) - \Pi(p_k)\| = 2(1 - \mathrm{e}^{-\lambda_k} - \lambda_k \mathrm{e}^{-\lambda_k}) \leqslant \lambda_k^2,$$

因而 $\|B - \Pi\| \leqslant \sum_{k=1}^n \lambda_k^2$.

2. 证明表达式 (9) 和 (10).

§13. 数理统计的基本定理

1. 数理统计与概率论的关系 在第一章 §7 中对于伯努利概型, 曾介绍过根据对随机变量的观测, 估计 "成功" 的概率, 以及为其置信区间的问题. 对于在一定意义上研究概率论的*反问题*的*数理统计*来说, 这些问题是典型的和基本的. 例如, 假如概率论的基本兴趣是, 对于给定的概率模型, 计算各种概率指标 (事件的概率, 随机元的概率分布及其各种特征, 等等), 则在数理统计中, 我们的兴趣在于: 如何根据统计资料 (以一定的可靠性) 揭示相应的概率模型, 使之在经验资料*统计性质*的框架内, 最好地与生成这些统计资料的随机机制的概率性质一致.

下面引用的 (格里汶科 [В. И. Гливенко], 坎泰利, 柯尔莫戈洛夫和斯米尔诺夫 [Н. В. Смирнов]) 结果, 理应称为*数理统计的基本定理*, 因为有关成果不但奠定了由统计资料提取 (关于被观测随机变量的分布函数的) 概率信息的*原则可能性*, 而且可以估计经验资料与各种不同的概率模型*拟合程度*.

2. 分布函数与经验分布函数的拟合定理 设 $\xi_1, \xi_2, \cdots, \xi_n$ 是定义在某一概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的*独立同分布*随机变量序列, 而 $F = F(x), x \in \mathbb{R}$, 是其共同的分布函数 $(F(x) = \mathbf{P}\{\xi_k \leqslant x\})$. 对于每一个 $N \geqslant 1$, 定义*经验分布函数*:

$$F_N(x;\omega) = \frac{1}{N}\sum_{k=1}^N I(\xi_k \leqslant x), \quad x \in \mathbb{R}. \tag{1}$$

根据大数定律 (§3 定理 2), 对于每一个 $x \in \mathbb{R}$, 有

$$F_N(x;\omega) \xrightarrow{\mathbf{P}} F(x), \quad N \to \infty, \tag{2}$$

即 $\mathbf{P}$- 依概率收敛.

由第四章 §3 的定理 1 和 2, 可见对于每一个 $x \in \mathbb{R}$, 当 $N \to \infty$ 时,依概率 1 收敛, 有

$$F_N(x;\omega) \to F(x), \quad (\mathbf{P} - \text{a.c.}). \tag{3}$$

这里有非常好的、更强的结果: (3) 式中的收敛关于 x 是一致的,

定理 (格里汶科和坎泰利) 在上述条件下, 随机变量

$$D_N(\omega) = \sup_{x\in R} |F_N(x;\omega) - F(x)| \tag{4}$$

依概率 1 收敛于 0:

$$\mathbf{P}\{\lim_N D_N(\omega) = 0\} = 1. \tag{5}$$

证明 设 $\mathbb{Q}$ 是 $\mathbb{R}$ 中有理数的集合. 显然

$$\sup_{r\in\mathbb{Q}} |F_N(r;\omega) - F(r)|$$

是随机变量. 由于

$$D_N(\omega) = \sup_{x\in\mathbb{R}} |F_N(x;\omega) - F(x)| = \sup_{r\in\mathbb{Q}} |F_N(r;\omega) - F(r)|,$$

可见 $D_N(\omega)$ 也是随机变量, 因此它有概率分布.

设整数 $M \geqslant 2$, 而 $k = 1, 2, \cdots, M-1$. 定义数列

$$x_{M,k} = \min\{x \in \mathbb{R} : k/M \leqslant F(x)\},$$

同时设 $x_{M,0} = -\infty, x_{M,M} = +\infty$.

假设 $x \in [x_{M,k}, x_{M,k+1})$, 而且 $[x_{M,k}, x_{M,k+1}) \neq \varnothing$. 那么, 显然

$$\begin{aligned}
&F_N(x;\omega) - F(x) \leqslant F_N(x_{M,k+1} - 0) - F(x_{M,k}) \\
&= [F_N(x_{M,k+1} - 0) - F(x_{M,k+1} - 0)] + [F(x_{M,k+1} - 0) - F(x_{M,k})] \\
&\leqslant F_N(x_{M,k+1} - 0) - F(x_{M,k+1} - 0) + \frac{1}{\nu}.
\end{aligned}$$

类似地, 再设 $x \in [x_{M,k}, x_{M,k+1})$, 有

$$F_N(x;\omega) - F(x) \geqslant F_N(x_{M,k};\omega) - F(x_{M,k}) - \frac{1}{M}.$$

因此, 对于每一个 $x \in \mathbb{R}$,

$$\begin{aligned}
&|F_N(x;\omega) - F(x)| \\
&\leqslant \max_{\substack{1\leqslant k\leqslant M-1 \\ 1\leqslant l\leqslant M-1}} \{|F_N(x_{M,k};\omega) - F(x_{M,k})|, |F_N(x_{M,l} - 0;\omega) - F(x_{M,l} - 0)|\} + \frac{1}{M},
\end{aligned}$$

于是,

$$\lim_{n\to\infty} \sup_x |F_N(x;\omega) - F(x)| \leqslant \frac{1}{M} \quad (\mathbf{P} - \text{a.c.}).$$

由于 M 的任意性, 由此可得结论 (5). □

格里汶科和坎泰利定理, 是数理统计的基本定理之一. 前面已经指出, 该定理奠定了由根据对独立同分布随机变量 $\xi_1, \xi_2, \cdots$ 的观测结果, 检验这些变量的分布函数是否恰好是函数 $F(x)$. 换句话说, 定理保障建立 "理论与试验" 一致的可能性.

3. 经验分布函数对分布函数的偏差 $D_N(\omega)$ 和 N 与 $F(x)$ 无关 尽管格里汶科和坎泰利定理有所指出的重要性, 然而它并未回答, 关于 "当 $N \to \infty$ 时, 偏差变量 $D_N(\omega)$ 收敛于 0 的*速度*" 问题. 因而, 不能推断, 独立观测结果与 "它 (观测结果) 具有假设的分布函数 $F = F(x), x \in \mathbb{R}$" 这一论断的*似然程度*.

由中心极限定理可见, 对于每一个*固定的* $x \in \mathbb{R}$, 有

$$\sqrt{N}[F_N(x;\omega) - F(x)] \xrightarrow{law} N(0, F(x)[1 - F(x)]), \tag{6}$$

说明随机变量 $\sqrt{N}[F_N(x;\omega) - F(x)]$ 收敛于均值为 0, 方差为 $F(x)[1 - F(x)]$ 的正态分布.

当然, 更重要的是得到 (一致) 统计量

$$D_N(\omega) = \sup_x |F_N(x;\omega) - F(x)|,$$

或相近的统计量

$$D_N^+(\omega) = \sup_x [F_N(x;\omega) - F(x)] \tag{7}$$

的极限分布.

在求这些统计量的极限分布时, 下面 (柯尔莫戈洛夫) 的研究结果是关键.

引理 1 *设 $\mathbb{F}$ 是连续型分布函数 $F = F(x)$ 类.*

对于每一个 $N \geqslant 1$ 和一切 $F \in \mathbb{F}$, 统计量 $D_N(\omega)$ 的概率分布是同一分布. 对于统计量 $D_N^+(\omega)$, 有类似的结果.

证明 设 $\eta_1, \eta_2, \cdots$ 是独立同分布随机变量序列, 且在 [0,1] 上同 ($U = U(x)$) 服从均匀分布: $\mathbf{P}\{\eta_1 \leqslant x\} = x, 0 \leqslant x \leqslant 1$.

由于对于连续函数 $F = F(x)$, 统计量 $\sup_x |F_N(x;\omega) - F(x)|$ 的分布, 与随机变量 $\sup_x |U_N(x;\omega) - U(x)|$ 的分布相同, 其中

$$U_N(x;\omega) = \frac{1}{N}\sum_{k=1}^{N} I(\eta_k(\omega) \leqslant x)$$

是随机变量 $\eta_1, \cdots, \eta_N$ 的经验分布函数, 由此可以得到引理的结论.

以 A 表示分布函数 $F = F(x)$ 是常数的区间 $I = [a,b], -\infty < a < b < \infty$, 的集合, 即 $\mathbf{P}\{\xi_1 \in I\} = 0$.

$$D_N(\omega) = \sup_{x\in\mathbb{R}} |F_N(x;\omega) - F(x)| = \sup_{x\in\overline{A}} |F_N(x;\omega) - F(x)|.$$

若引进随机变量 $\widetilde{\eta}_k = F(\xi_k)$ 及其经验分布函数

$$U_N(x;\omega) = \frac{1}{N}\sum_{k=1}^{N} I(\widetilde{\eta}_k(\omega) \leqslant x),$$

则对于任意 $x \in \overline{A}$,

$$U_N(F(x);\omega) = \frac{1}{N}\sum_{k=1}^{N} I(F(\xi_k(\omega)) \leqslant F(x)) = \frac{1}{N}\sum_{k=1}^{N} I(\xi_k(\omega) \leqslant x) = F_N(x;\omega),$$

因为对于这样的 $x, \{\omega : \xi_k(\omega) \leqslant x\} = \{\omega : F(\xi_k(\omega)) \leqslant F(x)\}$.

这样

$$\begin{aligned} D_N(\omega) &= \sup_{x\in\overline{A}} |F_N(x;\omega) - F(x)| = \sup_{x\in\overline{A}} |U_N(F(x);\omega) - F(x)| \\ &= \sup_{x\in\mathbb{R}} |U_N(F(x);\omega) - F(x)| = \sup_{y\in(0,1)} |U_N(y;\omega) - y| \\ &\overset{\mathbf{P}-\text{a.c.}}{=\!=\!=} \sup_{y\in[0,1]} |U_N(y;\omega) - y|, \end{aligned}$$

其中最后一个等式 ($\mathbf{P}$ − a.c.) 成立, 因为

$$\mathbf{P}\{\widetilde{\eta}_1 = 0\} = \mathbf{P}\{\widetilde{\eta}_1 = 1\} = 0. \tag{8}$$

现在证明, 随机变量 $\widetilde{\eta}_k$ (在 [0,1] 上) 均匀分布. 为此记

$$x(y) = \inf\{x \in \mathbb{R} : F(x) \geqslant y\}, \quad y \in (0,1).$$

那么, 有 $F(x(y)) = y, y \in (0,1)$, 且

$$\begin{aligned} \mathbf{P}\{\widetilde{\eta}_1 \leqslant y\} &= \mathbf{P}\{F(\xi_1) \leqslant y\} = \mathbf{P}\{F(\xi_1) \leqslant F(x(y))\} \\ &= \mathbf{P}\{\xi_1 \leqslant x(y)\} = F(x(y)) = y. \end{aligned}$$

由此连同 (8) 式就证明了, 随机变量 $\widetilde{\eta}_1$ (因此 $\widetilde{\eta}_2, \widetilde{\eta}_3, \cdots$ 也都) 在 $[0,1]$ 上均匀分布.

□

4. 统计量 $D_N(\omega)$ 和 $D_N^+(\omega)$ 的极限分布 上面的引理 1 表明, 为 (在对观测结果 $\xi_1, \xi_2, \cdots$ 的连续分布函数 $F = F(x)$ 类 $\mathbb{F}$ 中) 当 $N \to \infty$ 时求统计量 $D_N(\omega)$ 和 $D_N^+(\omega)$ 的极限分布, 只需假设 $\xi_1, \xi_2, \cdots$ 是独立同分布在 [0,1] 上均匀分布的随机变量序列.

固定某个 $N \geqslant 1$, 并且 (对于每一个 $\omega \in \Omega$) 将变量 $\xi_1, \xi_2, \cdots, \xi_N$ 按递增顺序排列所得的新变量 (**顺序统计量**) 记作 $\xi_1^{(N)}, \xi_2^{(N)}, \cdots, \xi_N^{(N)}$, 其中

$$\xi_1^{(N)} = \min(\xi_1, \xi_2, \cdots, \xi_N), \cdots, \xi_N^{(N)} = \max(\xi_1, \xi_2, \cdots, \xi_N).$$

(两个这样有序变量相等的概率等于 0.)

设

$$U_N(y;\omega) = \frac{1}{N}\sum_{k=1}^{N} I(\xi_k(\omega) \leqslant y), \tag{9}$$

则

$$D_N(\omega) = \max_{y \in \mathbb{R}} |U_N(y;\omega) - y|. \tag{10}$$

不难看出, 上式右侧只有在函数 $U_N(y;\omega)$ 的跳跃点 $\xi_1^{(N)}, \xi_2^{(N)}, \cdots, \xi_N^{(N)}$ 上, 才可以达到最大值. 因此,

$$D_N(\omega) = \max_{k \leqslant N} \left| \frac{k}{N} - \xi_k^{(N)} \right|. \tag{11}$$

由此可见, 为求统计量 $D_N(\omega)$ 的分布, 需要知道顺序统计量 $\xi_1^{(N)}, \xi_2^{(N)}, \cdots, \xi_N^{(N)}$ 的联合分布.

为求此分布, 考虑独立指数分布的随机变量序列: $\zeta_1, \zeta_2, \cdots$: $\mathbf{P}\{\zeta_k > x\} = \mathrm{e}^{-x}$, $x \geqslant 0$, 且设 $S_n = \zeta_1 + \cdots + \zeta_n, n \geqslant 1$.

引理 2 向量 $(\xi_1^{(N)}, \xi_2^{(N)}, \cdots, \xi_N^{(N)})$ 的联合分布, 等同于向量

$$\left(\frac{S_1}{S_{N+1}}, \frac{S_2}{S_{N+1}}, \cdots, \frac{S_N}{S_{N+1}} \right)$$

的联合分布.

证明 一方面, 对于 $0 \leqslant y_1 \leqslant \cdots \leqslant y_N \leqslant 1$, 有

$$\mathbf{P}\{\xi_1^{(N)} \in \mathrm{d}y_1, \cdots, \xi_N^{(N)} \in \mathrm{d}y_N\} = \sum \mathbf{P}\{\xi_{k_1} \in \mathrm{d}y_1, \cdots, \xi_{k_N} \in \mathrm{d}y_N\}, \tag{12}$$

其中对整数 $(1, \cdots, N)$ 的一切排列 $(k_1, \cdots, k_N)$ 求和.(在 (12) 式中及下面使用有些随意的 “微分” 书写形式, 不过不难赋予其确切的含义.)

在 (12) 式右侧的随机变量独立服从 [0,1] 上的均匀分布. 由于 $(1, \cdots, N)$ 的一切排列的总数为 $N!$, 可见 (12) 式右侧等于 $N!\mathrm{d}y_1 \cdots \mathrm{d}y_N$. 从而, 如果 $(y_1, \cdots, y_N) \in \Delta_N$, 其中 $\Delta_N = \{y_1 \in [0,1], \cdots, y_N \in [0,1] : 0 \leqslant y_1 \leqslant \cdots \leqslant y_N \leqslant 1\}$, 则

$$\mathbf{P}\{\xi_1^{(N)} \in \mathrm{d}y_1, \cdots, \xi_N^{(N)} \in \mathrm{d}y_N\} = N!\mathrm{d}y_1 \cdots \mathrm{d}y_N. \tag{13}$$

如果 $(y_1, \cdots, y_N) \notin \Delta_N$, 则

$$\mathbf{P}\{\xi_1^{(N)} \in \mathrm{d}y_1, \cdots, \xi_N^{(N)} \in \mathrm{d}y_N\} = 0. \tag{14}$$

另一方面, 现在证明同样的结果

$$\mathbf{P}\left\{ \frac{S_1}{S_{N+1}} \in \mathrm{d}y_1, \cdots, \frac{S_N}{S_{N+1}} \in \mathrm{d}y_N \right\} = \begin{cases} N!\mathrm{d}y_1 \cdots \mathrm{d}y_N, & 若\ (y_1, \cdots, y_N) \in \Delta_N, \\ 0, & 若\ (y_1, \cdots, y_N) \notin \Delta_N. \end{cases} \tag{15}$$

为此, 考虑随机变量 $S_1, \cdots, S_{N+1}$ 的联合分布.

如果 $(x_1,\cdots,x_{N+1})\in\Delta_{N+1}$, 则

$$\begin{aligned}&\mathbf{P}\{S_1\in \mathrm{d}x_1,S_2\in \mathrm{d}x_2,\cdots,S_{N+1}\in \mathrm{d}x_{N+1}\}\\ =&\mathbf{P}\{\zeta_1\in \mathrm{d}x_1,\zeta_2\in \mathrm{d}x_2-x_1,\cdots,\zeta_{N+1}\in \mathrm{d}x_{N+1}-x_N\}\\ =&\mathrm{e}^{-x_1}\mathrm{e}^{-(x_2-x_1)}\cdots\mathrm{e}^{-(x_{N+1}-x_N)}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_{N+1}\\ =&\mathrm{e}^{-x_{N+1}}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_{N+1}.\end{aligned}\tag{16}$$

如果 $(x_1,\cdots,x_{N+1})\notin\Delta_{N+1}$, 则

$$\mathbf{P}\{S_1\in \mathrm{d}x_1,S_2\in \mathrm{d}x_2,\cdots,S_{N+1}\in \mathrm{d}x_{N+1}\}=0.\tag{17}$$

因为 $S_{N+1}=\zeta_1+\cdots+\zeta_{N+1}$, 其中 $\zeta_1,\cdots,\zeta_{N+1}$ 是独立指数分布随机变量, 则 (练习题 3)

$$\mathbf{P}\{S_{N+1}\in \mathrm{d}x_{N+1}\}=\frac{x_{N+1}^N\mathrm{e}^{-x_{N+1}}}{N!}\mathrm{d}x_{N+1}.\tag{18}$$

由 (16) 和 (18) 式, 可见对于 $(x_1,\cdots,x_{N+1})\in\Delta_{N+1}$, 有

$$\mathbf{P}\{S_1\in \mathrm{d}x_1,\cdots,S_N\in \mathrm{d}x_N|S_{N+1}\in \mathrm{d}x_{N+1}\}=N!x_{N+1}^{-N}\mathrm{d}x_1\cdots\mathrm{d}x_N.\tag{19}$$

如果 $(x_1,\cdots,x_{N+1})\notin\Delta_{N+1}$, 则 (19) 式的左侧等于 0.

结果, 得

$$\begin{aligned}&\mathbf{P}\left\{\frac{S_1}{S_{N+1}}\in \mathrm{d}y_1,\cdots,\frac{S_N}{S_{N+1}}\in \mathrm{d}y_N\,\middle|\,S_{N+1}=x_{N+1}\right\}\\ =&\begin{cases}N!\mathrm{d}y_1\cdots\mathrm{d}y_N, & 若\ (y_1,\cdots,y_N)\in\Delta_N,\\ 0, & 若\ (y_1,\cdots,y_N)\notin\Delta_N.\end{cases}\end{aligned}$$

而由于右侧对 x_{N+1} 的独立性, 可得表示式 (15) 的正确性, 将其与 (13) 和 (14) 式比较, 便可证明引理 2 的结论. □

由引理 2 可见

$$\sqrt{N}D_N(\omega)\overset{d}{=}\sqrt{N}\max_{k\leqslant N}\left|\frac{S_k}{S_{N+1}}-\frac{k}{N}\right|,\tag{20}$$

其中 "$\overset{d}{=}$" 表示左, 右两侧随机变量的分布等同.

由 (20) 式, 有

$$\sqrt{N}D_N(\omega)\overset{d}{=}\frac{N}{S_{N+1}}\max_{k\leqslant N}\left|\frac{S_k-k}{\sqrt{N}}-\frac{k}{N}\frac{S_{N+1}-N}{\sqrt{N}}\right|.\tag{21}$$

对于研究当 $N\to\infty$ 时统计量 $\sqrt{N}D_N(\omega)$ 的极限性质, 关系式 (20) 非常方便.

由于 $N/S_{N+1}\to 1$ (P-a. c.), 则对于 $t\in[0,1]$, 设

$$X_t^{(N)}=\frac{S_{[Nt]}-[Nt]}{\sqrt{N}},$$

有

$$\lim_{N\to\infty}\mathbf{P}\{\sqrt{N}D_N(\omega)\leqslant y\}=\lim_{N\to\infty}\mathbf{P}\left\{\sup_{0\leqslant t\leqslant 1}|X_t^{(N)}-tX_1^{(N)}|\leqslant y\right\}.$$

这样, 关于统计量 $\sqrt{N}D_N(\omega)$ 的极限分布问题, 归结为当 $N\to\infty$ 时研究统计量

$$\sup_{0\leqslant t\leqslant 1}|X_t^{(N)}-tX_1^{(N)}|$$

的极限性质问题.

根据中心极限定理 (§4 定理 1), 对于每一固定的 $t\in[0,1]$, 随机变量 $X_t^{(N)}$ 的极限分布, 等同于随机变量 B_t 的极限分布 —— 参数为 $\mathbf{E}B_t=0$ 和 $\mathbf{E}B_t^2=t$ 的高斯分布.

实际上, 还可以得出更多结论. 具体地说, 考虑第二章 §13 引进的布朗运动 (维纳过程) $B=(B_t)_{0\leqslant t\leqslant 1}$. 这样的过程在第二章 §13, 是作为 $B_t=\mathbf{0},\mathbf{E}B_t=0$ 且协方差矩阵为 $\mathbf{E}B_sB_t=\min(s,t)$ 的高斯过程引进的. 于是, 结果表明, 根据第七章 §8 定理 1 的注 3, 随机变量 $(X_{t_1}^N,\cdots,X_{t_k}^N)$ 的联合分布 $P_{t_1,\cdots,t_k}^{(N)}$, 弱收敛于随机变量 $(B_{t_1},\cdots,B_{t_k})$ 的联合分布 $P_{t_1,\cdots,t_k}$. (关于弱收敛见 §1,2.) 自然想到, 统计量 $\sup_{0\leqslant t\leqslant 1}|X_t^{(N)}-tX_1^{(N)}|$ 及其分布, 收敛于统计量 $\sup_{0\leqslant t\leqslant 1}|B_t-tB_1|$ 及其分布.

一般说来, 有限维分布 $P_{t_1,\cdots,t_k}^{(N)}\to P_{t_1,\cdots,t_k},0\leqslant t_1<\cdots<t_k=1,k\geqslant 1$, 的弱收敛 (即按 §1 练习题 3 的记号, $P^{(N)}\xrightarrow{f}P$), 对于泛函

$$f(X^{(N)})=\sup_{0\leqslant t\leqslant 1}|X_t^{(N)}-tX_1^{(N)}|$$

的分布收敛于泛函

$$f(X)=\sup_{0\leqslant t\leqslant 1}|X_t-tX_1|$$

的分布还不充分, 其中 $X_t=B_t(0\leqslant t\leqslant 1)$ 是布朗运动.

不过, 对于所考虑的情形确实是这样, 这由以下的讨论可见.

过程 $X^{(N)}$ 的轨道属于空间 $D=D[0,1]$, 而过程 X 的轨道属于空间 $C=C[0,1]\subseteq D$ (见第二章 §2 第 6,7 小节). 在空间 D 中可以引进 "普罗霍罗夫度量" ρ (在 [5, 第三章] 中 ρ 记作 d_0, 在 [87, 第四章] 中 $\rho=\delta$), 度量空间 $(D,\mathscr{D},\rho)$ 关于 ρ 是完全可分空间. (第二章 §2 第 7 小节定义的, 以 "斯科罗霍德度量" d 为度量的空间 $(D,\mathscr{D},\rho)$ 仅是可分的, 而对于以后在第三章 §2 中的应用普罗霍罗夫定理还需要 "密度").

关于度量 ρ, 泛函 $f(x)=\sup_{0\leqslant t\leqslant 1}|x_t-tx_1|$ (显然与 $\sup_{0\leqslant t<1}|x_t-tx_1|$ 相等) 是连续的, 其中 $x=(x_t)_{0\leqslant t\leqslant 1}\in D$. 从而为证明收敛性 $f(X^{(N)})\xrightarrow{d}f(X)$ (即按分布收敛), 只需证明 (空间 $(D,\mathscr{D},\rho)$ 的) 弱收敛 $P^{(N)}\xrightarrow{w}P$ (练习题 2).

由普罗霍罗夫定理直接引进的, 证明这一收敛性的一般方法, 基于如下蕴涵关系 (见练习题 3, 以及 §1 练习题 3 的记号):

$$\Big([P^{(N)} \xrightarrow{f} P] \oplus [\{P^{(N)}\} \text{ 的密度 }] \oplus [\mathscr{K}_0(D) \text{ 是定义类}]\Big) \Rightarrow \Big(P^{(N)} \xrightarrow{w} P\Big). \tag{22}$$

(这里, $\mathscr{K}_0(D)$ 是柱集类;§1 第 5 小节.)

在所考虑的情形下, 是有限维分布的收敛 $P^{(N)} \xrightarrow{f} P$, 而柱集类 $\mathscr{K}_0(D)$ 确实是收敛的定义类 (第二章 §2 第 7 小节). 当然, 在运用蕴涵关系 (22) 时, 最困难的是验证所考虑的测度族 $\{P^{(N)}\}$ 是稠密的. 这样的验证基于包含在测度族定义中 (§2 的 (1) 式) 空间 D 的紧集的性质描述, 而这已经超出了本书的范围. (相应的证明, 参见 [5, 定理 15.2],[87, 第六章, 3.21]).

在这些讨论的最后我们指出, 还存在证明收敛 $f(X^{(N)}) \xrightarrow{d} f(X)$ 的如下途径. 间断过程 $X^{(N)}$ 可以用如下连续过程 $\widetilde{X}^{(N)}$ 逼近, 其中对于任意 $\varepsilon > 0$ 和一切 $\omega \in \Omega$, 以及充分大的 N, 有

$$\sup_{0\leqslant t\leqslant 1} |X_t^{(N)}(\omega) - \widetilde{X}_t^{(N)}(\omega)| \leqslant \varepsilon.$$

因此, 只需证明 $f(\widetilde{X}^{(N)}) \xrightarrow{d} f(X)$, 而这略微简单, 因为这时可以不涉及空间 D, 而涉及比较简单的空间 C. 在空间 C 中, 至少对于所考虑的情形, "稠密性" 准则容易验证 ([5, 第二章], [87, 第六章]).

5. 柯尔莫戈洛夫分布 设 $B_t^\circ = B_t - tB_1 (0 \leqslant t \leqslant 1)$. 这一过程是高斯过程, $B_0^\circ = 0, \mathbf{E}B_t^\circ = 0$, 而 $\mathbf{E}B_s^\circ B_t^\circ = \min(s,t) - st$. 在第二章 §13 第 7 小节, 此过程曾经称做*条件维纳过程*或*布朗桥*.

于是, 由以上的讨论得出这样的结论:

$$\lim_{N\to\infty} \mathbf{P}\{\sqrt{N}D_N(\omega) \leqslant y\} = \mathbf{P}\left\{\sup_{0\leqslant t\leqslant 1} |B_t^\circ| \leqslant y\right\}, \tag{23}$$

其中 $B^\circ = (B_t^\circ)_{0\leqslant t\leqslant 1}$ 是布朗桥.

由布朗运动的理论 (例如, 见 [5; 第二章, §11]) 知, 分布

$$K(y) = \mathbf{P}\left\{\sup_{0\leqslant t\leqslant 1} |B_t^\circ| \leqslant y\right\},$$

称做*柯尔莫戈洛夫分布*, 其中

$$K(y) = \sum_{k=-\infty}^{\infty} (-1)^k \mathrm{e}^{-2k^2y^2}, \quad y \geqslant 0. \tag{24}$$

这样, 有如下结果 (柯尔莫戈洛夫)

$$\lim_{N\to\infty} \mathbf{P}\{\sqrt{N}D_N(\omega) \leqslant y\} = K(y), \quad y \geqslant 0. \tag{25}$$

类似的讨论, 可得

$$\lim_{N\to\infty} \mathbf{P}\{\sqrt{N}D_N^+(\omega) \leqslant y\} = \mathbf{P}\left\{\sup_{0\leqslant t\leqslant 1} B_t^\circ \leqslant y\right\}, \tag{26}$$

统计量 $\sup_{0\leqslant t\leqslant 1} B_t^\circ$ 的分布, 比 $\sup_{0\leqslant t\leqslant 1}|B_t^\circ|$ 的分布简单. 统计量 $\sup_{0\leqslant t\leqslant 1} B_t^\circ$ 的分布为 (见 [5; 第二章, §11]):

$$\mathbf{P}\left\{\sup_{0\leqslant t\leqslant 1}|B_t^\circ| \leqslant y\right\} = 1 - \mathrm{e}^{-2y^2}, \quad y \geqslant 0. \tag{27}$$

有如下结果 (H. B. 斯米尔诺夫)

$$\lim_{N\to\infty} \mathbf{P}\{\sqrt{N}D_N^+(\omega) \leqslant y\} = 1 - \mathrm{e}^{-2y^2}, \quad y \geqslant 0. \tag{28}$$

6. 试验与实际的一致性准则 现在讨论, 由 (25) 式的结果 (其中 $K(y)$ 决定于 (24) 式), 可以建立试验与实际一致性的准则. 为此, 我们首先给出分布函数 $K(y)$ 的一张不大的数值表:

y	$K(y)$	y	$K(y)$	y	$K(y)$
		1.10	0.822 282	2.10	0.999 705
0.28	0.000 001	1.20	0.887 750	2.20	0.999 874
0.30	0.000 009	1.30	0.931 908	2.30	0.999 949
0.40	0.002 808	1.40	0.960 318	2.40	0.999 980
0.50	0.036 055	1.50	0.977 782	2.50	0.999 992 5
0.60	0.135 718	1.60	0.988 048	2.60	0.999 997 4
0.70	0.288 765	1.70	0.993 828	2.70	0.999 999 0
0.80	0.455 857	1.80	0.996 932	2.80	0.999 999 7
0.90	0.607 270	1.90	0.998 536	2.90	0.999 999 90
1.00	0.730 000	2.00	0.999 329	3.00	0.999 999 97

如果 N 充分大, 则可以认为 $K(y)$ 给出 $\mathbf{P}\{\sqrt{N}D_N(\omega) \leqslant y\}$ 值的很好的近似值.

自然, 如果按 $\xi_1(\omega), \cdots, \xi_N(\omega)$ 的经验值计算的变量 $\sqrt{N}D_N(\omega)$ 的值充分大, 则应当否定关于 "这些变量的 (假设的) 概率分布是 (连续型) 分布函数 $F = F(x)$" 的假设.

由上面表中的数值, 可以得出关于 "这样得到的结论的可靠程度" 的印象. 例如, 若 $\sqrt{N}D_N(\omega) > 1.80$, 则 (由于 $K(1.80) = 0.996\ 932$) 可以认为事件 $\{\sqrt{N}D_N(\omega) > 1.80\}$ 的概率大致等于 0.003 068(= 1.000 000 − 0.996 932). 假如认为 "具有这样小的概率 (= 0.003 068)" 的事件实际上出现的可能性很小, 则可以得出结论: 关于 "分布为 $\mathbf{P}\{\xi_1 \leqslant x\} = F(x)$" 的假设应当否定, 其中恰好是按函数 $F(x)$ 计算的 $\sqrt{N}D_N(\omega)$

的值. 而假如 $\sqrt{N}D_N(\omega) \leqslant 1.80$, 则 (根据大数定律) 可以认为在 1 000 000 次这样的情形下, 大致有 996 932 次 "经验与理论" 是一致的.

注 需要着重强调, 在运用基于 "柯尔莫戈洛夫分布和斯米尔诺夫分布" 的一致性准则时, (用于检验的) 分布函数 $F = F(x)$ 是完全确定的. 譬如, 只知道分布函数 $F = F(x)$ 属于分布函数 $G(x;\theta)$ 的某个参数族 $\mathbb{G} = \{G = G(x;\theta); \theta \in \Theta\}$, 而且 $G(x;\theta)$ 依赖于参数 $\theta \in \Theta$. 这时, 强求使用如下检验 "经验数据与符合真实分布函数 $F \in \mathbb{F}$" 途径: 首先根据 N 个观测值建立参数 θ 的估计, 然后计算

$$\widehat{D}_N(\omega) = \sqrt{N} \sup_{x \in \mathbb{R}} |F_N(x;\omega) - G(x; \widehat{\theta}_N(\omega))|,$$

并且像前面的例子一样采取相应的决定. 遗憾的是, 分布函数 $G_N(x;\widehat{\theta}_N(\omega))$ 是随机的, 而统计量 $\widehat{D}_N(\omega)$ 一般不再是柯尔莫戈洛夫分布.

因此, 在这种情形下, 关于如何检验假设:$F \in \mathbb{G}$, 参阅 [132].

7. 练习题

1. 证明 (18) 式.

2. 证明, 如果 (在空间 $(D, \mathscr{D}, \rho)$ 中) $P^{(N)} \xrightarrow{w} P$, 则 $f(X^{(N)}) \xrightarrow{d} f(X)$.

3. 证明蕴涵关系 (22).

图书文献资料

(第一章 ~ 第三章)

序言

在托德汉特 (I. Todhanter) 的专著 [68] 中, 陈述了拉普拉斯之前概率论的历史. 发表在《论文集》[45] 中的, 格涅坚科和舍伊宁 (О. В. Шейнин) 论文中, 介绍了从拉普拉斯到 19 世纪末的时期. 在斯蒂格列尔 (M. S. Stigler) 的书 [122] 中, 相当详细地介绍了 1900 年以前的概率论和数理统计史. 在迈斯特罗夫 (Д. Е. Майстров) 的书 [44] 中, 概率论的历史, 从概率论出现一直叙述到 20 世纪 30 年代. 在格涅坚科的教科书 [15] 中有概率论历史简介. 关于概率论许多术语的起源, 可参见亚历山大罗娃 (Н. В. Алексадрова) [2].

关于概率论的基本概念, 参见下列作者的书籍: 柯尔莫戈洛夫 [32], 格涅坚科 [15], 博罗夫科夫 [7], 格涅坚科和辛钦 [17], A. M. 亚格洛姆和 И. М. 亚格洛姆 (А. М. Яглом 和 И. М. Яглом) [83], 普罗霍罗夫和罗扎诺夫的手册 [56], 手册 [65] 以及译自英文的书: 费勒 [69], 奈曼 (Ю. Нейман) [51], 洛埃甫 (M. Loéve) [42], 杜布 (J. L. Doob) [20]. 这里还要指出包含大量概率论题目的梅沙尔金 (Л. Д. Мешалкин) 和基辅 "高等学校" 出版的习题集 [46] 和 [67]①.

在编写这一部教学参考书时, 作者使用了各种不同的文献. 在英文教材和教学指南中, 我们特别指出如下图书: 布雷曼 (L. Breiman)[8], 比林斯利 (P. Billingsley) [106], 阿什 (R. B. Ash) [80],[81], 阿什和加德钠 (M. F. Cadner) [82], 达雷特 (R. Durrett) [108]~[110], 兰珀蒂 (J. Lamperti) [37]. 本书的作者认为这些文献是成功传授知识的

①乌克兰文图书. —— 译者

典范.

关于概率论和数理统计方面的有益的参考资料, 读者可以利用《苏联大百科全书》,《苏联小百科全书》和《数学百科全书》① [121], 以及百科全书:《概率论和数理统计》[125].

于 1956 年创刊的杂志《概率论和数理统计》(出版社 “Наука”), 是俄罗斯出版的、概率论和数理统计方面的主要的刊物.

由全联盟科学与信息研究所 (莫斯科 ВИНИТИ)《文摘杂志》(《Реферативный журнал》), 刊登俄罗斯以及其他国家发表的 “概率论和数理统计方面” 论文的摘要.

有关概率统计大部分实际应用需要数值表, 可以使用博利舍夫 (Л. Н. Большев) 和斯米尔诺夫 (Н. В. Смирнов) 编著的《数理统计表》[6]. 在当前广泛使用电子计算机技术的条件下, 最好利用统计软件 MINITAB, SPSS, ··· 人们广泛采用 “Mathematica®” (见书 [126]).

第一章

§1. 关于概率模型的建立, 参见柯尔莫戈洛夫的论文 [31], 格涅坚科的书 [15]. 涉及 “按箱分配质点” 类型的问题大量的资料, 参见科尔钦 (В. Ф. Колчин), 塞伐斯契雅诺夫和奇斯佳科夫 (П. В. Чистяков) 的书 [34].

§2. 关于在统计物理中出现的概率模型 (特别, 一维伊金格 (Ezenger) 模型), 例如, 可以参见伊西哈尔 (А. Исихар) 的书 [25].

§3. 贝叶斯公式和定理, 是数理统计中所谓贝叶斯方法的基础. 例如, 可以参见德格鲁特 (M. H. deGroot) [18] 和扎克斯 (S. Zacks) [22].

§4. 关于随机变量及其概率特征, 可以参见习题集 [46] 和 [67].

§5. J. 伯努利引进的大数定律的组合证明, 例如可以参见 [69,T. I]. 关于大数定律的试验解释, 见柯尔莫戈洛夫的论文 [31].

§6. 有关局部和积分定理, 以及泊松定理中的偏差, 见博罗夫科夫 [7] 和普罗霍罗夫 [54].

§7. 这里叙述的内容, 是通过伯努利概型的例子, 演示某些基本概念和数理统计方法. 关于详细资料, 例如参见克拉默 (H. Cramér) [35] 和范德瓦尔登 (B. L. van der Waerden) [10].

§8. 通过考虑关于分割的条件概率和条件数学期望, 可以更好地掌握下面将要引进的, 较为复杂的关于 $\sigma-$ 代数的条件概率和条件数学期望的概念.

§9. 实际上, 这里所引进的 “破产问题” 的形式拉普拉斯就已经考虑过. 关于这一问题亦可参见文集 [45] 中格涅坚科和舍伊宁的论文. 在费勒的书 [69,T.I] 中有该

① 中译本:《数学百科全书》, 第一卷 ~ 第五卷, 科学出版社, 1994 — 2000 年, 北京. —— 译者

问题的丰富的资料.

§10. 这里的内容基本上是按费勒的书 [69] 叙述的. 在论文 [19] 中有关系式 (10) 和 (11) 的证明.

§11. 在杜布的书 [20] 中, 详细地阐述了鞅论. 例如可以在费勒的书 [69,T.I] 中找到 "表决" 定理的其他证明.

§12. 马尔可夫链大量资料包含在下列书中: 费勒 [69,T.I], 邓肯 [21], 邓肯和尤什克维奇 (А. П. Юшкевич), 钟开莱 [75],[120], 雷夫尤兹 (Д. Revuz), 凯麦尼 (J. G. Kemeny) 和斯奈尔 (J. L. Snell) [27], 萨雷姆萨科夫 (Т. А. Сарымсаков) [61], 西拉日季诺夫 (С. Х. Сиражцинов)[64]. 塞伐斯契雅诺夫 (Б. А. Севастьянов) 的专著 [62]《分枝过程》.

第二章

§1. 关于柯尔莫戈洛夫公理化体系见 [32].

§2. 关于 $\sigma-$ 代数的资料, 亦见下列图书: 柯尔莫戈洛夫和福明 (С. В. Фомин) [33], 奈维尤 (J. Neveu) [49], 布赖曼 [8], 阿什 [81].

§3. 卡拉泰奥多里定理的证明, 参见下列书: 洛埃甫 (M. Loéve) [42], 哈尔默斯 (P. R. Halmos) [70].

§4 ~ §5. 在哈尔默斯的书 [70] 中大量关于可测函数的资料.

§6. 亦见下列图书: 柯尔莫戈洛夫和福明 [33], 哈尔默斯 [70], 阿什 [81]. 在这些书中还有拉东 – 尼科迪姆定理的证明. 有时, 把

$$\mathbf{P}\{|\xi| \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}\xi^2}{\varepsilon^2}$$

称做切比雪夫不等式, 而把

$$\mathbf{P}\{|\xi| \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}|\xi|^r}{\varepsilon^r}, \quad r > 0,$$

称做马尔可夫不等式.

§7. 关于 $\sigma-$ 代数的条件概率和条件数学期望的定义, 是柯尔莫戈洛夫引进的 [32]. 在布赖曼 [8] 和阿什 [81] 的书中, 有所考虑问题的广泛资料.

§8. 亦见下列图书: 博罗夫科夫 [7], 阿什 [81], 克拉默 [35], 格涅坚科 [15].

§9. 柯尔莫戈洛夫关于 "具有给定有限维分布过程的存在性" 的定理见他的专著 [32]. 关于图尔恰 (I. Tulcea) 定理, 亦见奈维尤 [49], 阿什 [81]. 这里进行的证明遵循 [81].

§10 ~ 11. 亦见柯尔莫戈洛夫和福明 [33], 阿什 [81], 杜布 [20], 洛埃甫 [42].

§12. 特征函数理论在许多图书上都有, 例如: 格涅坚科 [15], 柯尔莫戈洛夫和格涅坚科 [16], 拉曼钱德兰 (B. Lamanchandran) [57]. 所介绍的矩和半不变量的联系按照列昂诺夫 (В. П. Леонов) 和施利亚耶夫的论文 [40].

§13. 亦见下列图书: 易卜拉给莫夫 (И. А. Ибрагимов), 罗扎诺夫 [24], 布赖曼 [8], 利普彩尔 (Р. Ш. Липцер) 和施利亚耶夫 [41], 格里米特 (G. R. Grimmit) 和斯特扎克 (D. R. Stirzaker) [105], 兰珀蒂 [37].

第三章

§1. 关于概率测度和分布的弱收敛, 在下列书中有详细的陈述: 格涅坚科和柯尔莫戈洛夫 [16], 比林斯利 [5].

§2. 普罗霍罗夫定理在他的论文 [55] 中.

§3. 格涅坚科和柯尔莫戈洛夫的专著 [16] 的内容中, 有概率论的极限定理证明中的特征函数方法. 亦见比林斯利 [5]. 练习题第 2 题包括: J. 伯努利大数定律, 泊松大数定律, 假设满足条件: $\xi_1, \xi_2, \cdots$ 独立, 有 0 和 1 两个可能值, 但一般有不同的分布, 即 $\mathbf{P}\{\xi_i = 1\} = p_i, \mathbf{P}\{\xi_i = 0\} = 1 - p_i, i \geqslant 1$.

§4. 这里, 在林德伯格条件成立时, 进行独立随机变量之和的、中心极限定理的传统证明. 对照 [16],[92].

§5. 在没有极限可忽略的条件下, 中心极限定理成立的条件问题, 曾引起 P. 列维的注意. 在佐洛塔廖夫 (В. М. Золотарёв) 专著 [88] 中详细地介绍了, 极限定理理论在*非经典提法*下的现状. 罗塔里 (В. И. Ротарь) [96] 给出了定理 1 的提法和证明.

§6. 这一节利用了下列图书的资料: 格涅坚科和柯尔莫戈洛夫的专著 [16], 阿什 [81], 彼得罗夫 (В. В. Петров)[53],[92].

§7. 莱维 – 普罗霍罗夫度量, 是普罗霍罗夫在其著名的著作 [55] 中引进的, 定义在度量空间上的、测度之弱收敛的可度量性的结果也属于普罗霍罗夫. 关于度量 $\|P - \widetilde{P}\|^*_{BL}$, 参见达德利 (R. M. Dudley) [85] 和波拉德 (D. Pollard) 的书 [93].

§8. 定理 1 属于斯科罗霍德 (А. В. Скороход). 关于 "一个概率空间方法" 的有益资料, 可以在下列文献在找到: 博罗夫科夫 1999 年教学参考书, 专著波拉德 [33].

§9 ~ 10. 我们列举包含 "关于所涉及问题" 大量资料的一系列图书: 扎克德 (J. Jacod) 和施利亚耶夫 [87], 卡姆 [L. Le Cam] [89], 格林伍得 (P. E. Greenword) 和施利亚耶夫 [84], 莱斯 (F. Liese) 和威伊达 (I. Vajda) [90].

§11. 在彼得罗夫的专著 [92] 中, 有关于中心极限定理收敛速度估计的大量资料. 所引用的贝里和埃森定理的证明, 包含在格涅坚科和柯尔莫戈洛夫的专著 [16] 中.

§12. 证明借鉴普雷斯曼 (I. L. Pressman) 的文章 [94].

§13. 关于数理统计的基本定理的补充资料, 参见 [8],[35],[58],[106],[107].

参考文献

[1] Александров П. С. Введение в общую теорию множеств и функций. — М.: Гостехиздат, 1948.

[2] Александрова Н. В. Математические термины. — М.: Высшая школа, 1978.

[3] Бернштейн С. Н. О работах П. Л. Чебышева по теории вероятностей//Научное наследие П. Л. Чебышева. Вып. 1:Математика. — 1945. — С. 59 — 60.

[4] Бернштейн С. Н. Теория вероятностей. — 4-е изд. — М.: Гостехиздат, 1946.

[5] Биллингсли П. Сходимость вероятностных мер. — М.: Наука, 1977.

[6] Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. — 3-е изд. — М.: Наука, 1983.

[7] Боровков А. А. Теория вероятностей. — 3-е изд. — М.:УРСС, 1999.

[8] Брейман (Breiman L.). Probability. — Reading, MA: Addison-Wesley, 1968.

[9] Вальд А. Последовательный анализ. — М.: Физматгиз, 1960.

[10] Ван дер Варден Б. Л. Математическая статистика. — М.: ИЛ, 1960.

[11] Вентцель А. Д. Курс теории случайных процессов. — М.: Наука, 1975.

[12] Гарсиа (Garsia A. M.). A simple proof of E. Hopf's maximal ergodic theorem // Journal of Mathematics and Mechanics. — 1965. — V.14, № 3. — P. 381 — 382.

[13] Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. — М.: Наука, 1977.

[14] Гихман И. И., Скороход А. В. Теория случайных процессов: В 3 т. — М.: Наука, 1971 — 1975. 中译本:《随机过程论》, 第一卷 (邓永录等译), 第二卷 (周概容译), 1986.

[15] Гнеденко Б. В. Курс теории вероятностей. —6-е изд. — М.: Наука, 1998. 中译本:《概率论教程》(丁寿田), (第三版), 高等教育出版社, 1961.

[16] Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. — М.; Л.: Гостехиздат, 1949. 中译本:《相互独立随机变量之和的极限分布》(王寿仁译), 科学出版社, 1950.

[17] Гнеденко Б. В., Хинчин А. Я. Элементарное введение в теорию вероятностей. — 9-е изд. — М.: Наука, 1982.

[18] Де Гроот М. Оптимальные статистические решения. — М.: Мир, 1974.

[19] Дохерти (Doherty M.). An amusing proof in fluctuation theory// Combinatorial Mathematics, III: Proceedings of the Third Australian Conference, Univ. Queensland, St. Lucia 1974. — Berlin etc.: Springer-Verlag, 1975. — P. 101—104. — (Lecture Notes in Mathematics; V.452.)

[20] Дуб Дж. Л. Вероятностные процессы. — М.: ИЛ, 1956.

[21] Дынкин Е. Б. Марковские процессы. — М.: Физматгиз, 1963.

[22] Закс Ш. Теория статистических выводов. — М.: Мир, 1975.

[23] Ибрагимов И. А., Линник Ю. В. Независимые и стационарно связанные величины. — М.: Наука, 1965.

[24] Ибрагимов И. А., Розанов Ю. А. Гауссовские случайные процессы. — М.: Наука, 1970.

[25] Исихара А. Статистическая физика. — М.: Мир, 1973.

[26] Кабанов Ю. М., Липцер Р. Ш., Ширяев А. Н. К вопросу об абсолютной непрерывности и сингулярности вероятностных мер// Математический сборник. — 1977, — Т. 104. № 2. — С. 227 — 247.

[27] Кемени Дж., Снелл Дж. конечные цепи Маркова. — М.: Наука, 1970.

[28] Колмогоров А. Н. Цепи Маркова со счетным числом возможных состояний//Бюллетень МГУ. — 1937. — Т.1, № 3. — С. 1 — 16.

[29] Колмогоров А. Н. Стационарные последовательности в гильбертовском пространстве//Бюллетень МГУ. — 1941. — Т. 2, № 6. — С. 1 — 40.

[30] Колмогоров А. Н. Роль русской науки в развитии теории вероятностей//Ученые записки МГУ. — 1947. — Вып. 91. — С. 53 — 64. 中译本:《概率论》(见《数学, 它的内容、方法和意义》(卷2)), 科学出版社, 1961.

[31] Колмогоров А. Н. Теория вероятностей//Математика, ее содержание, методы и значение.—М.: Изд-во АН СССР, 1956. — Т. II. — С. 252 — 284.

[32] Колмогоров А. Н. Основные понятия теории вероятностей. — М.; Л.: ОНТИ, 1936; 2-е изд. М.: Наука, 1974; 3-е изд. М.: Фазис, 1998.

[33] Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — 6-е изд. —М.: Наука, 1989. 中译本:《概率论的基本概念》(丁寿田译), 商务印书馆, 1952.

[34] Колчин В. Ф., Севастьянов Б. А., Чистяков В. П. Случайные размещения. — М.: Наука, 1976.

[35] Крамер Г. Математические методы статистики. — 2-е изд. — М.: Мир, 1976. 中译本:《统计学数学方法》(魏棕舒译), 上海科技出版社, 1983.

[36] Кубилюс И. Вероятностные методы в теории чисел. — Вильнюс: Гос. изд-во полит. и науч. лит. ЛитССР, 1959.

[37] Ламперти Дж. Вероятность. — М.: Наука, 1973.

[38] Ламперти (Lamperti J.). Stochastic Processes. — Aarhus Univ., 1974. — (Lecture Notes Series; № 38).

[39] Ленгляр (Lenglart E.). Relation de domination entre deux processus// Annales de l'Institut H. Poincaré Sect. B. (N. S.). — 1977. — V.13, № 2. — P. 171 — 179.

[40] Леонов В. П., Ширяев А. Н. К технике вычисления семиинвариантов// Теория вероятностей и ее применения. — 1959. — Т. IV, вып. 2. — С. 342 — 355.

[41] Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. — М.: Наука, 1974. 中译本:《随机过程统计》.

[42] Лоэв М. Теория вероятностей. — М.: ИЛ, 1962. 中译本:《概率论及其应用》, 上册, (梁文骐译), 科学出版社, 1966.

[43] Марков А. А. Исчисление вероятностей. — 3-е изд. — СПб., 1913.

[44] Майстров Д. Е. Теория вероятностей (исторический очерк). — М.: Наука, 1967.

[45] Математика XIX века/ Под ред. А. Н. Колмогорова и А. П. Юшкевича. — М.: Наука, 1978.

[46] Мешалкин Л. Д. Сборник задач по теории вероятностей. — М.: Изд-во МГУ, 1963. 中译本:《概率论习题集》(盛骤等译), 高等教育出版社, 1984.

[47] Мейер (Meyer P.-A.). Martingales and Stochastic Integrals. I. — Berlin etc.: Springer-Verlag, 1972. — (Lecture Notes in Mathematics; V. 284).

[48] Мейер П.-А. Вероятность и потенциалы. — М.: Мир, 1973.

[49] Неве Ж. Математические основы теории вероятностей. — М.: Мир, 1969.

[50] Невё (Neveu J.). Discrete-Parameter Martingales.—Amsterdam etc.: North-Holland, 1975.

[51] Нейман Ю. Вводный курс теории вероятностей и математической статистики. — М.: Наука, 1968.

[52] Новиков А. А. Об оценках и асимптотическом поведении вероятностей непересечения подвижных границ суммами независимых случайных величин//Известия АН СССР. Серия математическая. — 1980. — Т. 40, вып. 4. — С. 868 — 885.

[53] Петров В. В. Суммы независимых случайных величин. — М.: Наука, 1972

[54] Прохоров Ю. В. Асимптотическое поведение биномиального распределения// Успехи математических наук. — 1953. — Т. VIII, вып. 3(55). — С. 135—142.

[55] Прохоров Ю. В. Сходимость случайных процессов и предельные теоремы теории вероятностей// Теория вероятностей и ее применения. — 1956. — Т. I, вып. 2. — С. 177—238.

[56] Прохоров Ю. В., Розанов Ю. А. Теория вероятностей. — 2-е изд. — М.: Наука, 1973.

[57] Рамачандран Б. Теория характеристических функций. — М.: Наука, 1975.

[58] Реньи (Rényi A.) Probability Theory. — Amsterdam: North-Holland, 1970.

[59] Роббинс Г., Сигмунд Д., Чао И. Теория оптимальных правил остановки. — М.: Наука, 1977.

[60] Розанов Ю. А. Стационарные случайные процессы. — М.: Физматгиз, 1963.

[61] Сарымсаков Т. А. Основы теории процессов Маркова. — М.: Гостехиздат, 1954.

[62] Севастьянов Б. А. Ветвящиеся процессы. — М.: Наука, 1971.

[63] Синай Я. Г. Введение в эргодическую теорию. — Ереван: Изд-во Ереван. ун-та, 1973.

[64] Сиражднинов С. Х. Предельные теоремы для однородных цепей Маркова. — Ташкент: Изд-во АН УзССР, 1955.

[65] Справочник по теории вероятностей и математической статистике/ Под ред. В. С. Королюка. — Киев: Наукова думка, 1978.

[66] Стоут (Stout W. F.). Almost Sure Convergence. — New York etc.: Academic Press, 1974.

[67] Теорія имовірностей. — Київ: Вища школа, 1976.

[68] Тодхантер (Todhunter I.). A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace. — London: Macmillan, 1865.

[69] Феллер В. Введение в теорию вероятностей и ее приложения: В 2-х т. — М.: Мир, 1984. 中译本:《概率论导引及其应用》(胡迪鹤译), 人民邮电出版社, 2006.

[70] Халмош П. Теория меры. — М.: ИЛ, 1953. 中译本:《测度论》(王建华译), 科学出版社.

[71] Хеннан Э. Анализ временных рядов. — М.: Наука, 1964.

[72] Хеннан Э. Многомерные временные ряды. — М.: Мир, 1974.

[73] Чао, Тейчер (Chow Y. S., Teicher H.). Probability Theory. Independence, Interchangeability, Martingales. — 3rd ed. — New York: Springer-Verlag, 1997.

[74] Чебышев П. Л. Теория вероятностей: Лекции акад. П. Л. Чебышева, читанные в 1879, 1880 гг./ Издано А. Н. Крыловым по записи А. М. Ляпунова. — М.; Л., 1936.

[75] Чжун Кай-лай. Однородные цепи Маркова. — М.: Мир, 1964.

[76] Ширяев А. Н. Случайные процессы. — М.: Изд-во МГУ, 1972.

[77] Ширяев А. Н. Вероятность, статистика, случайные процессы: В 2-х т. — М.: Изд-во МГУ, 1973 — 1974.

[78] Ширяев А. Н. Статистический последовательный анализ. — 2-е изд. — М.: Наука, 1976.

[79] Энгельберт, Ширяев (Engelbert H.-J., Shiryaev A. N.). On the sets of convergence of generalized submartingales // Stochastics. — 1979. — V. 2, № 3. — P. 155 — 166.

[80] Эш (Ash R. B.). Basic Probability Theory. — New York etc.: Wiley, 1970.

[81] Эш (Ash R. B.). Real Analysis and Probability.—New York etc.: Academic Press, 1972.

[82] Эш, Гарднер (Ash R. B., Gardner M. F.). Topics in Stochastic Processes. — New York etc.: Academic Press, 1975.

[83] Яглом А. М., Яглом И. М. Вероятность и информация. — 3-е изд. — М.: Наука, 1973. 中译本:《概率与信息》, 科学出版社.

[84] Гринвуд, Ширяев (Greenwood P. E., Shiryayev A. N.). Contiguity and the Statistical Invariance Principle. — London: Gordon & Breach, 1985.

[85] Дадли (Dudley R. M.)Distances of probability measures and random variables// Annals of Mathematical Statistics. — 1968. — V.39, № 5. — P. 1563 — 1572.

[86] Дакуна-Кастелль, Дюфло (Dacunha-Castelle D., Duflo M.). Probabilités et statistiques: 1,2. — Paris: Masson. — 1: Problèmes à temps fixe. — 1982; — 2: Problèmes à temps mobile. — 1983. — Перев. на англ. яз.: Probability and Statistics: V. I, II. — Berlin etc.: Springer-Verlag, 1986.

[87] Жакод Ж., Ширяев А. Н. Предельные теоремы для случайных процессов: В 2-х т. — М.: Физматлит, 1994.

[88] Золотарев В. М. Современная Теория суммирования независимых случайных величин. — М.: Наука, 1986.

[89] Ле Кам (Le Cam L.). Asymptotic Methods in Statistical Decision Theory. — Berlin etc.: Springer-Verlag, 1986.

[90] Лизе, Вайда (Liese F., Vajda I.). Convex Statistical Distances. — Leipzig: Teubner, 1987.

[91] Липцер Р. Ш., Ширяев А. Н. Теория мартиигалов. — М.: Наука, 1986

[92] Петров В. В. Предельные теоремы для сумм независимых случайных величин. — М.: Наука, 1987.

[93] Поллард (Pollard D.). Convergence of Stochastic Processes. — Berlin etc.: Springer-Verlag, 1984.

[94] Пресман Э. Л. О сближении по вариации распределения суммы независимых бернуллиевских величин с пуассоновским законом// Теория вероятностей и ее применения. — 1985. — Т. XXX, вып. 2. — С. 391 — 396.

[95] Розанов Ю. А. Теория вероятностей, случайные процессы и математическая статистика. — М.: Наука, 1985.

[96] Ротарь В. И. К обобщению теоремы Линдеберга — Феллера// Математические заметки.— 1975. — Т.18, вып. 1. — С. 129 — 135.

[97] Севастьянов Б. А. Курс теории вероятностей и математической статистики. — М.: Наука, 1982.

[98] Ширяев (Shiryayev A. N.) Probability. — 2nd ed. — Berlin etc.: Springer-Verlag, 1995.

[99] Ширяев (Shirjayev A. N.) Wahrscheinlichkeit. — Berlin: VEB Deutscher Verlag der Wissenschaften, 1988.

[100] Ширяев А. Н. Основы стохастической финансовой математики: В 2-х т. — М.: ФАЗИС, 1998.

[101] Фёллмер, Проттер, Ширяев (Föllmer H., Protter Ph., Shiryaev A. N.). Quadratic covariation and an extension of Itô's formula//Bernoulli. — 1995. — V.1, № 1/2. — P. 149 — 170.

[102] Дынкин Е. Б., Юшкевич А. А. Теоремы и задачи о процессах Маркова. — М.: Наука, 1967.

[103] Гнеденко, Колмогоров (Gnedenko B. V., Kolmogorov A. N). Limit Distributions for Sums of Independent Random Variables. — Reading, MA, etc.: Addison-Wesley, 1954.

[104] Боровков А. А. Эргодичность и устойчивость случайных процессов. — М.: УРСС, 1999.

[105] Гриммет, Стирзакер (Grimmet G. R., Stirzaker D. R.). Probability and Random Processes. — Oxford: Clarendon Press, 1993.

[106] Биллингсли (Billingsley P.). Probability and Measure. — 3rd ed. — New York: Wiley, 1995.

[107] Боровков А. А. Математическая статистика. — М.: Наука, 1984.

[108] Дарретт (Durrett R.). Probability: Theory and Examples. — Pacific Grove, CA: Wadsworth & Brooks/Cole, 1991.

[109] Дарретт (Durrett R.). Stochastic Calculus. — Boca Raton, FL: CRC Press, 1996.

[110] Дарретт (Durrett R.). Brownian Motion and Martingales in Analysis. — Belmont, CA: Wadsworth International Group, 1984.

[111] Калленберг (Kallenberg O.). Foundations of Modern Probability. — 2nd ed. — New York: Springer-Verlag, 2002.

[112] Карлин, Тейлор (Karlin S., Taylor H. M.). A First Course in Stochastic Processes. — 2nd ed. — New York etc.: Academic Press, 1975.

[113] Кашин Б. С., Саакян А. А. Ортогональные ряды. — 2-е изд. — М.: АФЦ, 1999.

[114] Жакод, Проттер (Jacod J., Protter Ph.). Probability Essentials. — Berlin etc.: Springer-Verlag, 2000.

[115] Нётс (Neuts V. F.) Probability. — Boston, MA: Allyn & Bacon, 1973.

[116] Плато (Plato J.). Creating Modern Probability. — Cambridge: Cambridge Univ. Press, 1998.

[117] Ревюз Д. Цепи Маркова. — М.: РФФИ, 1997.

[118] Уильямс (Williams D.). Probability with Martingales. — Cambridge: Cambridge Univ. Press, 1991.

[119] Холл, Хейде (Hall P., Heyde C. C.). Martingale Limit Theory and Its Applications. — New York etc.: Academic Press, 1980.

[120] Чжун Кай-Лай (Chung Kai Lai). Elementary Probability Theory with Stochastic Processes. — 3rd ed. — Berlin etc.: Springer-Verlag, 1979. 中译本:《初等概率论和随机过程》(吕乃刚等译), 人民教育出版社.

[121] Математическая энциклопедия: В 5 т./Гл. ред. И. М. Виноградов. — М.: Советская энциклопедия, 1977 — 1985. 中译本:《数学百科全书》, 第一卷 ~ 第五卷, 科学出版社, 1984 — 2000.

[122] Стиглер (Stigler S. M.).The History of Statistics: The Measurement of Uncertainty Before 1900. — Cambridge: Belknap Press of Harvard Univ. Press, 1986.

[123] Гербер Х. Математика страхования жизни. — М.: Мир, 1995.

[124] Эренфесты П. и Т. (Ehrenfest P., Ehrenfest T.). Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem//Physikalische Zeitschrift. — 1907. — V.8. — P. 311 — 314.

[125] Теория вероятностей и математическая статистика: энциклопедия/ Гл. ред. Ю. В. Прохоров. — М.: Большая Российская энциклопедия, 1999.

[126] Вольфрам (Wolfram S.). The Mathematica® Book. — 4th ed. — Champaign; Cambridge: Wolfram Media; Cambridge Univ. Press, 1999.

[127] Дуб (Doob J. L.). What is a martingale?//The American Mathematical Monthly. — 1971. — V.78. — P. 451 — 463.

[128] Синай Я. Г. Курс теории вероятностей. — М.: Изд-во МГУ, 1985. — 2-е изд., 1986.

[129] Синай (Sinaĭ Ya. G.). Topics in Ergodic Theory. — Princeton, NJ: Princeton Univ. Press, 1999. — (Princeton Mathematical Series; V. 44.)

[130] Вальтерс (Walters P.) An Introduction to Ergodic Theory. — New York etc.: Springer-Verlag, 1982.

[131] Булинский А. В., Ширяев А. Н. Теория случайных процессов. — М.: Физматлит, 2003.

[132] Хмаладзе Э. В. Мартингальный подход в теории непараметрических критериев согласия// Теория вероятностей и ее применения. — 1981. — Т. XXVI. вып. 2. — С. 246 — 265.

[133] Гамильтон (Hamilton J. B.). Time Series Analysis. — Princeton, NJ: Princeton Univ. Press, 1994.

[134] Бернулли Я. О законе больших чисел. — Ч. 4: Искусство предположений. — М.: Наука, 1986.

[135] Лукач Е. Характеристические функции. — М.: Наука, 1979.

[136] Хренников (Khrennikov A.). Interpretations of Probability. — Utrecht: VSP, 1999.

名词索引

(汉语拼音为序)

C

H

J

M

N

P

Q

Z

人名表

(汉语拼音为序)

A

阿什	R. B. Ash	Р. Эш
埃尔米特	Ch. Hermite	Ш. Эрмит
埃森	C. G. Esseen	К. Г. Эссеен
爱因斯坦	A. Einstein	А. Эйнштейн

B

巴拿赫	S. Banach	С. Банах
邦弗尔罗尼	Bonferroni	Бонферрони
鲍斯	S. N. Bose	Ш. Бозе
贝尔	A. G. Bell	А. Г. Белл
贝尔	R. L. Baire	Р. Л. Бэр
贝里	A. C. Berry	А. С. Берри
贝塞尔	F. W. Bessel	Ф. В. Бессель
贝叶斯	T. Bayes	Т. Байес
比林斯利	P. Billingsley	П. Биллингсли
彼得罗夫	W. W. Petrov	В. В. Петров
毕达哥拉斯	Pythagoras	Пифагор
波拉德	D. Pollard	Д. Пполлард
波利亚	G. Polya	Д. Пойа

伯恩斯坦	S. N. Bernstein	Н Бернштейн
伯克霍夫	G. D. Birkhoff	Дж. Д. Биркгоф
伯努利	J. Benoulli	Я. Бернулли
泊松	S. D. Poisson	С. Д. Пуассон
博赫纳	S. Bochner	С. Бохнер
博雷尔 (波莱尔)	E. Borel	Э. Борель
博利舍夫	L. N. Bolishev	Л. Н. Большев
博罗夫科夫	A. A. Borowkov	А. А. Боровков
布尔	G. Boole	Дж. Буль
布耳兹曼	L. Boltzmann	Л. Больцман
布莱克韦尔	D. H. Blackwell	Д. Блэкуэлл
布赖曼	L. Breiman	Л. Брейман
布朗	E. T. Brown	Э. Т. Броун
布雷	J. R. Bray	Я. Р. Брэй
布洛赫	A. Bloch	А. Блох
布尼亚科夫斯基	A. J. Buniakowsky	А. Я. Буняковский
布西	R. S. Busey	Р. С. Вьюсу

C~D

查普曼	D. G. Chapman	Д. Г. Чепман
达布	J. G. Darboux	Ж. Г. Дарбу
达德利	R. M. Dudley	Р. Дадли
达雷特	R. Durrett	Р. Даррет
德格鲁特	M. H. deGroot	М. Де Гроот
邓肯	E. B. Dynkin	И. Б. Дынкин
狄拉克	P. A. M. Dirac	П. А. М. Дирак
狄利克雷	P. G. L. Dirichlet	П. Г. А. Дирихле
笛卡尔	R. Descartes	Р. Декарт
棣莫弗	A. Dé Moivre	А. Дé Муавр
杜布	J. L. Doob	Дж. Л. Дуб

F

法图	F. Fatou	П. фату
范德瓦尔登	B. L. van der Waerden	Б. Л. Ван дер Варден
范基	Fan Ky	Ки Фан
费勒	W. Feller	В. Феллер
费马	P. Fermat	П. Ферма
费希尔	R. A. Fisher	Р. А. Фишер

弗雷歇	M. Fréchet	Ф. Фреше
弗米	E. Fermi	Э. Ферми
福明	S. W. Fomin	С. В. Фомин
傅比尼	G. Fubini	Г. Фубини
傅里叶	J. B. J. Fourier	Ж. Б. Ж. Фурье

G

冈贝尔	E. J. Gumbel	Э. Гумбель
高斯	G. F. Gauss	К. Ф. Гаусс
哥塞特	W. S. Gosset	В. С. Госсет
克拉默	G. P. Gram	Г. П. Грам
格里米特	G. R. Grimmit	Дж. Гриммет
格里汶科	W. I. Glivenko	В. И. Гливенко
格林伍得	P. E. Greenword	П. Е. Гринвуд
格涅坚科	B. V. Gnedenko	Б. В. Гнеденко

H

哈恩	H. Hahn	Г. Хан 460
哈尔	A. Haar	А. Хаар
哈尔默斯	P. R. Halmos	П. Халмош
海林格	E. Helinger	Э. Хеллингер
海涅	H. E. Heine	Г. Э. Нейне
赫尔德	O. L. Hölder	О. Л. Гёльдер
赫利	E. Helly	Э. Хелли
惠更斯	Ch. Huyghens	Х. Гюйгенс
霍普夫	H. Hopf	Х. Хопф
霍奇	W. W. D. Hodge	У. В. Д. Ходе

J

加德钠	M. F. Cadner	М. Гарднер
加尔西亚	A. M. Garsia	А. М. Гарсиа

K

卡尔达诺	G. Cardanno	Дж. Кардано
卡尔莱曼	T. Carleman	Т. Карлеман
卡库塔尼	S. Kakutani	С. Какутани
卡拉泰奥多里	C. Carathéodory	К. Каратеодори
卡姆	L. Le Cam	Л. Ле Кам
凯麦尼	J. G. Kemeny	Дж. Кемени

坎泰利	F. P. Cantelli	Ф. П. Кантелли
康托尔	G. Cantor	Г. Контор
柯尔莫戈洛夫	A. N. Kolmogorov	А. Н. Колмогоров
柯西	A. L. Cauchy	О. Л. Коши
科尔钦	W. F. Kerchen	В. Ф. Колчин
克拉默	H. Cramer	Г. Крамер
克罗内克	L. Kronecker	Л. Кронекер
库尔贝克	S. Kullback	С. Кульбак

L

拉奥	C. R. Rao	С. Р. Рао
拉德马赫	H. Rademacher	Г. А. Радемахер
拉东	J. Radon	Дж. Радон
拉格朗日	J. L. Lagrange	Ж. Л. Лагранж
拉曼钱德兰	B. Lamanchandran	Б. Раманчадран
拉普拉斯	P. S. Laplace	П. С. Лаплас
莱斯	F. Liese	Ф. Лизе
兰珀蒂	J. Lamperti	Дж. Ламперти
勒贝格	H. L. Lebesgue	А. Л. Лебег
雷夫尤兹	Д. Revuz	Д. Ревюз
黎曼	G. F. B. Riemann	Г. Ф. Б. Риман
李雅普诺夫	A. M. Lyapunov	А. М. Ляпунов
里斯	F. Riesz	Ф. Рисс
利普彩尔	R. S. Lipchail	Р. Ш. Липцер
利普希茨	R. O. S. Lipschitz	Р. Липшиц
列昂诺夫	W. P. Leonov	В. П. Леонов
列维	P. P. Lévy	П. П. Леви
林德伯格	J. W. Linderberg	Дж. У. Линдеберг
洛埃甫	M. Loéve	М. Лоэв
洛必达	L' Hospital	Г. Лопиталь
罗塔里	W. E. Rotari	В. И. Ротарь
罗扎诺夫	Yu. A. Rozanov	Ю. А. Роэанов

M

马尔可夫	A. A. Markov	А. А. Марков
马钦凯维奇	J. Matcinkiewicz	Й. Марцинкевич
马哈拉诺比斯	P. S. Mahalanobis	П. С. Махаланобис
迈斯特罗夫	D. I. Mastrov	Д. Е. Майстров

麦克米兰	B. McMillan	Б. Макмиллан
麦克斯韦	J. C. Maxwell	Д. К. Максвелл
曼	H. B. Mann	Х. Б. Манн
梅沙尔金	L. D. Mesharlkin	Л. Д. Мешалкин
米泽斯	R. Mises	Р. Миэесу
闵可夫斯基	H. Minkowski	Г. Минковский
默瑟	J. Mercer	Ж. Мерсер

N

奈曼	Yv. Neyman	Ю. Нейман
奈维尤	J. Neveu	Р. Невё
尼科迪姆	O. M. Nikodym	О. М. Никодим

O

欧几里得	Euclid	Евклид
欧拉	L. Euler	Л. Эйлер

P

帕利	W. Pauli	В. Паули
帕乔利	L. Papcioli	Л. Пачоли
帕塞瓦尔	M. A. Parseval	М. А. Пасеваль
帕斯卡	B. Pascal	Б. Паскаль
庞加莱	J. H. Poincaré	Ж. Ан. Пуанкаре
蒲丰 (布丰)	G. L. L. Buffon	Ж. Л. Л. Бюффон
普拉特	Pratt	Пратт
普雷斯曼	I. L. Pressman	Э. Л. Пресман
普罗霍罗夫	Yu. B. Prokhorov	Ю. В. Прохров

Q

奇斯佳科夫	W. P. Qisjiakv	В. П. Чистяков
切比雪夫	P. L. Chebyshev	П. Л. Чебьшёв
切萨罗	E. Cesàro	Э. Чеэаро 335

S

萨雷姆萨科夫	T. A. Saramsakov	Т. А. Сарымсаков
塞维治	I. R. Sevage	И. Р. Сэвидж
沙利耶	C. L. Charlier	К. Л. Шарлье
绍德尔	A. Schauder	А. Шаудер
舍伊宁	O. W. Sherining	О. В. Шейнин

施利亚耶夫	A. N. Shiryaev	А. Н. Ширяев
施密特	E. Schimidt	Э. Шмидт
施瓦茨	L. Schwarz	Л. Шварц
斯梯格列尔	S. M. Stigler	С. М. Сти
斯蒂尔切斯	T. J. Stieltjes	Т. И. Стилтьес
斯捷克洛夫	W. A. Steclov	В. А. Стеклов
斯科罗霍德	A. V. Skorokhod	А. В. Скороход
斯莱皮恩	P. Slepian	П. Слепян 394
斯鲁斯基	E. Slutsky	Е. Е. Слуцкий
斯米尔诺夫	N. V. Smirnov	Н. В. Смирнов
斯奈尔	J. L. Snell	Дж. Снелл
斯皮策	F. Spitzer	Ф. Спцер
斯特林	J. Stirling	Дж. Стирлинг
斯特扎克	D. R. Stirzaker	Д. Стирэакер
斯蒂格列尔	M. S. Stigler	С. Стиглер
斯通	M. H. Stone	М. Г. Стоун

T

塔尔塔利亚	N. Tataylia	Н. Тарталья
图尔恰	I. Tulcea	И. Тулча
托德汉特	I. Todhanter	Э. Тодхантер
托德亨特	I. Todhunter	И. Тодхатер

W

威伊达	I. Vajda	И. Вайда
韦布尔	W. Weibull	В. Вейбулл
维尔斯特拉斯	K. T. W. Weirstrass	К. Т. В.
维纳	N. Weiner	Н. Винер
沃尔德	A. Wald	А. Вальд
沃尔夫	R. Wolf	Р. Вольф
乌沙科夫	W. G. Ushakoff	В. Г. Ушаков
伍拉姆	Woollam	Улам

X

西拉日季诺夫	S. H. Sealarijinov	С. Х. Сираждинов
希尔伯特	D. Hilbert	Д. Гильберт
希奈	Y. G. Sinaĭ	Я. Г. Синай

谢瓦斯契亚诺夫	B. A. Sevastyanov	Б. А. Севастьянов
辛钦	A. J. Khintchine	А. Я. Хичин
休伊特	E. Hewitt	Э. Хьюитт

Y

雅可比	G. G. J. Jaccbi	К. Г. Я. Якоби
亚当斯	J. G. Adams	Д. К. Адамс
亚格洛姆	A. M. Jaglom	А. М. Яглом
亚格洛姆	E. M. Jaglom	И. М. Яглом
亚历山大罗娃	N. W. Alexanderova	Н. В. Алексадрова
延森	I. L. Iensen	И. Л. Иенсен
伊金格	Ezenger	Изинг
伊藤	K. Itô	К. Ито
伊西哈尔		А. Исихар
易卜拉给莫夫	E. A. Ibrageimov	И. А. Ибрагимов
尤什克维奇	A. P. Yushikaivici	А. П. Ю шкевич

Z

扎克德	J. Jacod	Ж. Жакод
扎克斯	S. Zacks	Ш. Эакс
钟开莱	Kai Lai Chung	Чжун Кай -лай
祖布科夫	A. M. Zobkov	А. М. Эубков
佐洛塔廖夫	W. M. Zolotaileoff	В. М. Эолотарёв

常用数学符号

$\mathbb{R}=(-\infty,\infty)$ —— 实数的集合, 实直线, 一维欧几里得空间
$\mathbb{R}_+=[0,\infty)$
$\overline{\mathbb{R}}=[-\infty,\infty]$ —— 扩充实直线: $\overline{\mathbb{R}}=\mathbb{R}\bigcup\{-\infty\}\bigcup\{\infty\}$
$\overline{\mathbb{R}}_+=[0,\infty]$
$\mathbb{Q}$ —— 有理数的集合
$\mathbb{Q}_+=\mathbb{Q}\bigcap\mathbb{R}_+$
$\mathbb{R}^{\mathrm{d}}$ ——d 维欧几里得空间
$\mathbb{N}$ —— 自然数: $\{0,1,2,\cdots\}$ 或 $\{1,2,\cdots\}$
$\mathbb{Z}$ —— 整数的集合: $\{0,\pm1,\pm2,\cdots\}$
$\mathbb{C}$ —— 复数的集合

$$(a,b)=\{x\in\overline{\mathbb{R}}:a<x<b\},[a,b]=\{x\in\overline{\mathbb{R}}:a\leqslant x\leqslant b\}$$

$$(a,b]=\{x\in\overline{\mathbb{R}}:a<x\leqslant b\},[a,b)=\{x\in\overline{\mathbb{R}}:a\leqslant x<b\}$$

$\inf X$ —— 集合 $X\subseteq\overline{\mathbb{R}}$ 的下界
$\sup X$ —— 集合 $X\subseteq\overline{\mathbb{R}}$ 的上界
$\inf\limits_{n\geqslant m}x_n$ —— 集合 $X=\{x_m,x_{m+1},\cdots\}$ 的下界
$\sup\limits_{n\geqslant m}x_n$ —— 集合 $X=\{x_m,x_{m+1},\cdots\}$ 的上界

如果$x_n\in\overline{\mathbb{R}},n\geqslant 1$, 则

$$\liminf_{n\to\infty}x_n=\underline{\lim}_n x_n\equiv\sup_{m\geqslant 1}\inf_{n\geqslant m}x_n,\quad \limsup_{n\to\infty}x_n=\overline{\lim}_n x_n\equiv\inf_{m\geqslant 1}\sup_{n\geqslant m}x_n,$$

$$\lim x_n=x\Leftrightarrow\underline{\lim}x_n=\overline{\lim}x_n=x\Leftrightarrow\underline{\lim}x_n\geqslant x\geqslant\overline{\lim}x_n.$$

对 于 实 数

$x^+ = \max(x,0), x^- = -\min(x,0)$

$x^{\oplus} = \begin{cases} x^{-1}, & 若\ x \neq 0, \\ 0, & 若\ x = 0 \end{cases}$

$x \vee y = \max(x,y), x \wedge y = \min(x,y)$

$[x]$ 或 $\lfloor x \rfloor$ —— 不大于 x 的最大整数

$\lceil x \rceil$ —— 大于或等于 x 的最小整数

sign x —— 实数 x 的符号:

$$\text{sign } x = \begin{cases} 1, & 若\ x > 0, \\ 0, & 若\ x = 0, \\ -1, & 若\ x < 0 \end{cases}$$

(有时, 当 $x \geqslant 0$ 时, 设 sign $= 1$; 当 $x < 0$ 时, 设 sign $= -1$)

$x_n \to x$, 其中 $n \in \{1, 2, \cdots\}$, 表示 $\lim\limits_n x_n = x$

$x_n \uparrow$ 表示 $x_1 \leqslant x_2 \leqslant \cdots$; $x_n \uparrow x$ 表示 $x_n \uparrow$ 且 $\lim\limits_n x_n = x$

$x_n \downarrow$ 表示 $x_1 \geqslant x_2 \geqslant \cdots$; $x_n \downarrow x$ 表示 $x_n \downarrow$ 且 $\lim\limits_n x_n = x$

对于复数 $z = a + ib$, 其中 $a, b \in \mathbb{R}$, 而 $i = \sqrt{-1}$ 是虚单位

$\overline{z} = a - ib$ —— z 的共轭复数

$|z|$ —— z 的模 $(= \sqrt{a^2 + b^2})$

Re z 和 Im z —— z 的实部和虚部: Re $z = a$, Im $z = b$

对于 $d-$ 维欧几里得空间 $\mathbb{R}^{\mathrm{d}}$

$|x|$ —— $x = (x_1, \cdots, x_d)$ 的欧几里得范数, 即 $\sqrt{x_1^2 + \cdots + x_d^2}$

$x \cdot y$ 或 (x, y) —— $x = (x_1, \cdots, x_d)$ 和 $y = (y_1, \cdots, y_d)$ 的数量积, 即 $x_1 y_1 + \cdots + x_d y_d$

集 合 论

$A_n \uparrow$ 表示 $A_1 \subseteq A_2 \subseteq \cdots$; $A_n \uparrow A$ 表示 $A_n \uparrow$ 且 $\bigcup A_n = A$

$A_n \downarrow$ 表示 $A_1 \supseteq A_2 \supseteq \cdots$; $A_n \downarrow A$ 表示 $A_n \downarrow$ 且 $\bigcap A_n = A$

$\limsup A_n$, 或 $\overline{\lim} A_n$, 或 $\bigcap\limits_{m \geqslant 1} \left(\bigcup\limits_{n \geqslant m} A_n \right)$ —— 属于无限多个集合 $A_n (n \geqslant 1)$ 的点的集合

$\liminf A_n$, 或 $\underline{\lim} A_n$, 或 $\bigcup\limits_{m \geqslant 1} \left(\bigcap\limits_{n \geqslant m} A_n \right)$ —— 属于所有 (仅可能有限个 A_n 除外) 集

合 $A_n(n \geqslant 1)$ 的点的集合

I_A 或 $I(A)$ —— 集合 A 的示性函数

$\{\cdots\}$ —— 集合

数 学 符 号

$\ll$ —— 绝对连续

$\sim$ —— 等价

$\perp$ —— 垂直

$f = o(g) - \lim\left(\frac{f}{g}\right) = 0$

$f = O(g) - \limsup\left|\frac{f}{g}\right| < \infty$

$f \sim g - \lim\left(\frac{f}{g}\right) = 1$

$f \asymp g$ —— 比值 $\frac{f}{g}$ 自下与 0 分离, 且自上与 ∞ 分离

$f \circ g$ —— 复合

$f * g$ —— 卷积

［General Information］
书名＝概率　第１卷　修订和补充第三版
作者＝（俄罗斯）Ａ．Ｈ．施利亚耶夫著
页数＝４５７
ＳＳ号＝１１８５２５０９
出版日期＝２００７．７